TEC-0039

Multivariate Spectral Analysis to Extract Materials from Multispectral Data

Robert S. Rand
Donald A. Davis

September 1993

U.S. Army Corps of Engineers
Topographic Engineering Center
Fort Belvoir, Virginia 22060-5546

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible] [illegible] [illegible]

[illegible]

[illegible] [illegible]

[illegible] [illegible]

[illegible]

[illegible] [illegible]

[illegible]

[illegible] [illegible]

[illegible] [illegible]

[illegible] [illegible] [illegible] [illegible]

[illegible] [illegible]

ILLUSTRATIONS

TABLES

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible] with additional [illegible] as well as [illegible] and physically relevant [illegible]

[illegible] higher spatial and/or [illegible] the data resulting [illegible]

[illegible] provide motivation [illegible] to perform [illegible]

2.0 APPROACH

2.1 Selection of Appropriate Algorithms

Numerous classification algorithms were considered as candidate methods for extracting natural and manmade features. These included the parametric supervised classifiers such as the Bayesian discriminant and Mahalanobis distance classifiers; non-parametric supervised classifiers such as the simple Euclidean minimum distance, and error correction techniques such as the Ho-Kashyap and Widrow-Hoff methods; as well as unsupervised clustering techniques such as K-Means and the ISODATA methods. Because these methods are commonly documented, knowledge about them is assumed, and details are only brought into the discussion as needed.[5] Mathematical descriptions of the selected algorithms are given for reference and for the sake of being precise about what is actually being tested.

Past experience, along with some theoretical considerations, led investigators to exclude clustering methods from the current effort. Such methods are perhaps best suited for sorting pixels in a non-homogeneous training class into a small number of homogeneous ones, as discussed in Section 2.2.1. However, clustering on an image containing anything but the simplest of scenes should be avoided. During an effort conducted during the Persian Gulf War that was directed at detecting oil against a water background, the ISODATA/ISOCLASS method was found to give unstable results.[6] In particular, two Landsat TM images containing almost identical scenes were clustered using the same ISODATA process and running parameters. One of the resultant class map images displayed very impressive results that were in fact judged better than the results produced from the Bayesian discriminant and Euclidean minimum distance methods; however, the second image produced results that were nonsense and totally useless for delineating oil. KMEANS is a simpler algorithm which is an alternative; however, this clustering method requires a priori knowledge of the number of clusters. Both methods are, of course, nonparametric.

From a mathematical viewpoint, the disadvantage in using ISODATA/ISOCLASS is that finding a unique global solution cannot be guaranteed. This clustering technique may settle into a local rather than global solution (the minimized value of its objective function is not a global minimum). The local solution generally depends on the initial starting estimates for the seed clusters and specifying different seed points for the initial clusters can produce different classification outputs. The differences may or may not be significant, but nevertheless a unique solution can never be guaranteed. In the case of the Persian Gulf study, the results from the second image apparently settled into such a local minimum, and this solution did not correspond to the reality of the ground features within the scene.

The error-correction procedures (nonparametric) were not considered because of the desire to ultimately use a rejection criterion for pixels that do not match a training class or that correspond to a mixture of classes (the need for this rejection capability is discussed below). From a theoretical viewpoint, the most appealing approach to invoking a rejection statistic is to work within the framework of a parametric model. Although a parametric-based rejection statistic could be computed separately, it seemed more appropriate to use a parametric model throughout this stage of

---

[5] Charles W. Therrien. *Decision Estimation and Classification*. New York, NY: John Wiley & Sons, 1989.
Sing-Tze Bow. *Pattern Recognition - Applications to Large Data-Set Problems*. New York, NY: Marcel Dekker, Inc., 1984.

[6] Robert Rand, Donald Davis, M.B. Satterwhite and John Anderson. *Methods of Monitoring the Persian Gulf Oil Spill Using Digital and Hardcopy Multiband Data*. Fort Belvoir, VA: U.S. Army Topographic Engineer Center, TEC-0014, August 1992.

classification. Also, some limited experience with the Widrow-Hoff method [illegible] solution (although guaranteed to converge) could be rather slow to converge.

Neural networks, such as training by back propagation, are a relatively new approach [illegible] promising; however, they are also computationally very intensive and would have [illegible] much effort to implement and study, given the resources available. If [illegible] conventional multivariate methods are found to be significant and cannot be [illegible] network approach may be a promising alternative.

Therefore, based on the above arguments, the focus of this effort would be on three [illegible] Euclidean minimum distance, the Bayesian discriminant, and Mahalanobis distance. The Euclidean minimum distance was included as an alternative simple method to use as a [illegible] to the other more complex methods. Given that this method is perhaps the simplest of classification methods, one can examine the degree of improvement or degradation [illegible] invoking more complex mathematical models. All these classifiers make use of a [illegible] $g_i(\mathbf{x})$ to select a class for each observation vector $\mathbf{x}$. Given that there are [illegible] decision rule is to choose the class ω_i which corresponds to the maximum of [illegible]

In an effort to resolve the mixed-pixel problem, a linear mixing model was [illegible] basic method is built on the statistical linear modeling approach as is commonly done in [illegible] analysis. Spectral endmembers (usually, pure pixels) are defined as the independent [illegible] variables, and the mixed pixel of interest is defined as the dependent variable. [illegible] recently been proposed by certain researchers for broad-band and narrow-band [illegible] data (see footnotes 3 and 4 Section 1.0). As will be discussed, we feel this method [illegible] approached with caution. However, a couple of constraints can be placed on the [illegible] help screen the number of physically allowable combinations, and [illegible] conform to what is physically expected from a linear mixing phenomenon. By [illegible] physical constraints that will be discussed, we attempt to overcome the [illegible] basic model and avoid misusing the linear regression method. This approach, [illegible] constraints, is proposed in Section 2.1.5 and later analyzed by experiment [illegible]

2.1.1 Euclidean Minimum Distance Classifier.

The Euclidean minimum distance classifier is simple and computationally [illegible] classifier, meaning that the decision surfaces are hyperplanes. The decision function [illegible]

$$g_i(\mathbf{x}) = -r_i^2(\mathbf{x}) = -(\mathbf{x} - \mu_i)^t(\mathbf{x} - \mu_i)$$

where $\mathbf{x}$ is the n dimensional pixel vector being classified, and μ_i is the n dimensional [illegible] for class ω_i. Notice the maximum $g_i(\mathbf{x})$ corresponds to the minimum squared distance. The function $g_i(\mathbf{x})$ is evaluated for each class, and the pixel is assigned to the class with the maximum value of $g_i(\mathbf{x})$.

This method is most appropriate when the components of a vector are independent [illegible] variances. In our case of broad-band spectral data, this means the bands should be [illegible] and have equal variance. Of course, it is commonly known that neither is the [illegible] dimensionality of the image data is quite high, or the classes are spectrally well [illegible] unlikely that such linear surfaces will be adequate to segment the imagery into the [illegible]

2.1.2 Bayesian Classifier

The Bayesian classifier is a quadratic algorithm that generates hyperquadric decision surfaces (i.e. hyperplanes, hyperspheres, hyperellipsoids, hyperparaboloids). Accordingly, it is also more complex and computationally slower. From a statistical point of view, the algorithm is attractive because it weights the variables, and it accounts for correlation among them. Under the assumption that class data belong to multivariate normal populations, the method is optimal in the sense that it minimizes the probability of classification error. The multivariate normal (MVN) assumption allows the distributional properties of each class to be completely specified by a mean vector and covariance matrix. Unfortunately, violations of the MVN assumption (quite common in practice) and difficulties in estimating the class covariance matrices can potentially lead to poor performance.

The conditional probability function for a multivariate normal random vector $\mathbf{x} \sim MVN(\mu, \Sigma)$ belonging to class ω_i is

$$f_{X|W}(\mathbf{x}|\,\omega_i) = \frac{1}{(2\pi)^{n/2}|\Sigma_i|^{1/2}} \exp\left[\, -\tfrac{1}{2} * (\mathbf{x} - \mu_i)^t \Sigma_i^{-1} (\mathbf{x} - \mu_i) \,\right]$$

where Σ_i is the covariance matrix for class ω_i, and n is the dimension of each pixel vector $\mathbf{x}$ and each mean vector μ_i.

The Bayes classifier appeals to the well-known Bayes Theorem and then uses the logarithm of the *a posterior* probability $f_{W|X}(\omega_i|\mathbf{x}) = f_{X|W}(\mathbf{x}|\,\omega_i) * P(\omega_i)$ as the definition of the Bayes discriminant function:

$$g_i(\mathbf{x}) = -\tfrac{1}{2} * (\mathbf{x} - \mu_i)^t \Sigma_i^{-1} (\mathbf{x} - \mu_i) - \tfrac{1}{2} \log |\Sigma_i| + \log P(\omega_i) - \tfrac{n}{2} \log 2\pi$$

During this study, the *a priori* probabilities $P(\omega_i)$ are set equal and do not contribute to the decision. Since the last term is a constant that also does not contribute to the decision, the Bayes discriminant function used in this study is

$$g_i(\mathbf{x}) = -\tfrac{1}{2} * (\mathbf{x} - \mu_i)^t \Sigma_i^{-1} (\mathbf{x} - \mu_i) - \tfrac{1}{2} \log |\Sigma_i|$$

In obtaining good performance, the MVN assumption seems to be more critical for the quadratic classifiers (such as Bayes) than it is for the linear ones.[7] One reason for this is that the mathematical properties of the true decision regions are well behaved for MVN prototype (training) distributions and can be defined by positive definite quadratic forms. For example, the regions are defined by conic sections in the bivariate case (two multispectral bands). The classification region for a particular class might be the interior of an ellipse or the region between two hyperbolas.[8] In general, a quadratic function will define the regions; however, it is not necessarily a positive

---

[7] Richard A. Johnson, Dean W. Wichern. *Applied Multivariate Statistics*. 2nd Edition, Englewood Cliffs, NJ: Prentice-Hall, 1988, p 493 and p513.

[8] T. W. Anderson. *An Introduction to Multivariate Statistical Analysis*. 2nd Edition, New York, NY: John Wiley & Sons, 1984, p235.

definite quadratic form. In this case, the Bayes classifier as defined is no longer optimal since the model is only an approximation.

Poor performance can result from difficulties in estimating class covariance matrices. Such difficulties can result from either insufficient variation in a sample (attributable to lack of feature variation and/or quantization effects) or inappropriately high variation (attributable to non-homogeneous samples and/or outliers). This issue is discussed further in Section 2.2.2.

However, a major contributor to poor performance is mixed pixels comprised of more than one feature. If a mixture comprised mostly of a predominant material is used as a training sample, the MVN assumption is almost certainly violated. The covariance estimate for the predominant class will also be too high and therefore may give the class distribution too high a spread (ideal training classes should have low variance/covariance to reduce the overlap between classes). If the training data are constrained to pure pixels, mixtures in the remaining image data can skew the corresponding pixel vector intensities toward the wrong class, resulting in misclassifications.

If the classes of interest are well separated, violation of the MVN assumption usually does not generate poor performance, so long as the distribution is reasonably symmetric. The major culprit seems to be mixed pixels.

2.1.3 Mahalanobis Distance Classifier

The Mahalanobis distance classifier is similar in complexity to the Bayesian, except that rather than making the decision based on the probability function, it simply uses the squared Mahalanobis distance from the pixel of concern and each of the prototype class centers. Like the Bayesian method, it is a quadratic classifier. The discriminant function is minus the squared Mahalanobis distance:

$$g_i(x) = -d_i^2(x) = -(x - m_i)^T \Sigma_i^{-1} (x - m_i)$$

As with the minimum Euclidean distance, notice the maximum $g_i(x)$ corresponds to the minimum squared Mahalanobis distance $d_i^2(x)$. Also notice that this function is identical to the dominating quadratic Bayesian term.

2.1.4 Mahalanobis Distance Threshold Analysis as a Rejection Criterion

Based on a multivariate normal assumption for the random vector x, the distribution of the squared Mahalanobis distance random variable $d_i^2(x) = (x - m_i)^T \Sigma_i^{-1} (x - m_i)$ is chi-squared with n degrees of freedom (n is the dimension of the vector x).[3] That is

$$d_i^2 \sim \chi^2(n)$$

The property that $d_i^2(x)$ is a chi-squared variable can be used to provide a rejection criterion. Given a significance value α, the classification should be rejected if

[3] Richard A. Johnson, Dean W. Wichern, Applied Multivariate Statistical Analysis, 2nd Edition, Englewood Cliffs, NJ: Prentice Hall, 1988.

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

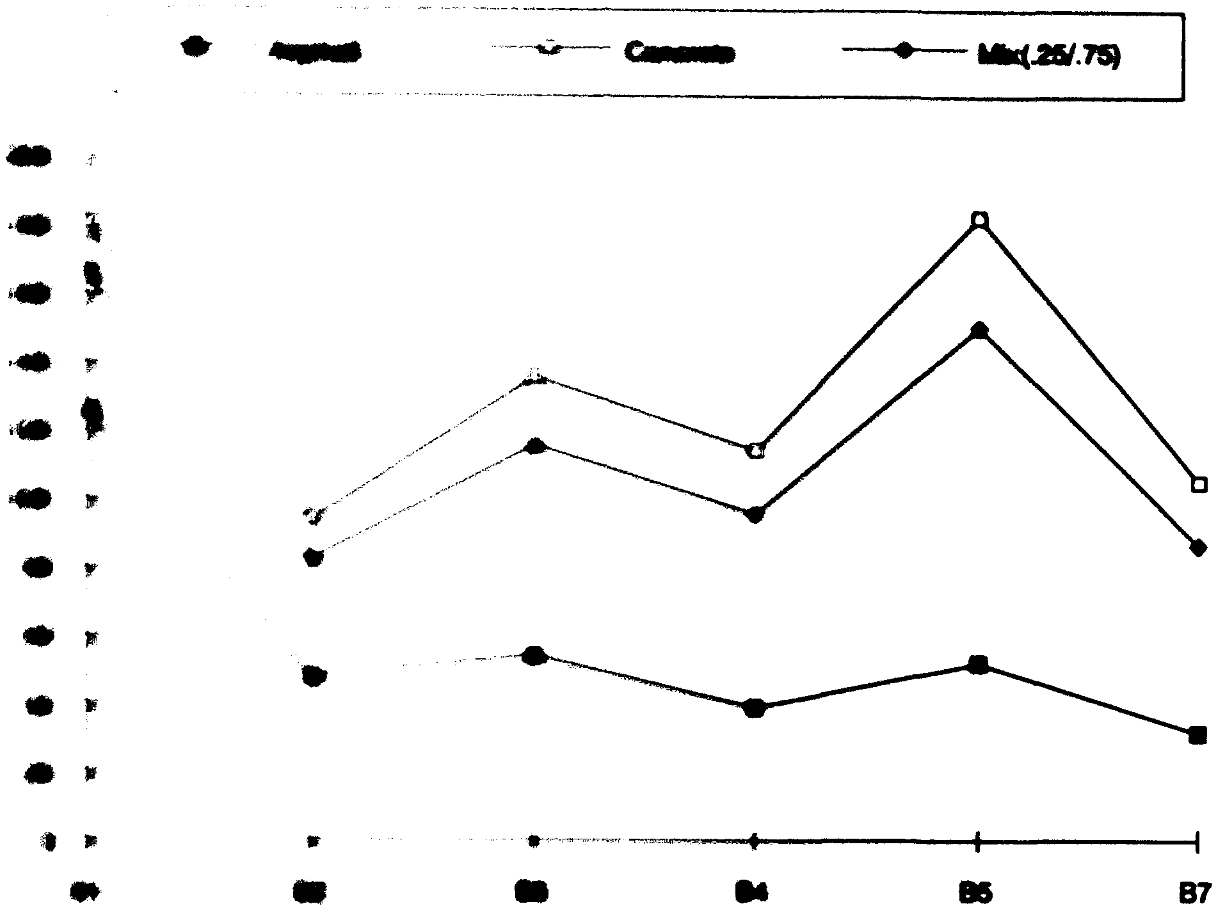


Figure 1. Predicted Linear Mixture of Asphalt and Concrete.

Overall Utility of the Method. From a theoretical point of view, the basic method has a number of potentially severe limitations, and it is reasonable to question whether the method is worth pursuing. This method might have some utility for handling mixtures, but only if the limitations and ways to handle them can be characterized. By imposing the two physical constraints mentioned above, we attempt to overcome the inherent limitations of the basic model and avoid [illegible] the linear regression method.

Another concern is that the entire spectrum is weighted equally in this model. Observing the spectra for various materials, one can quickly notice there are wide swings in certain regions of the spectra for some materials, but not for others (see Section 4.1).

2.2 Basic Issues

There are three basic issues that need to be addressed regarding techniques to extract natural and manmade materials from broad-band imagery: the optimal selection of training classes, improving the performance of conventional algorithms, and handling mixtures of materials.

2.2.1 Optimal Selection of Training Classes

The three supervised methods require training data to define prototype classes. It is quite conceivable that the performance of these classifiers will vary significantly, depending on the skill of an analyst to define appropriate prototype classes. Not only must such training classes be spectrally separable from each other, they must be representative of the features in the rest of the scene. There are the issues of whether to choose a large or small number of classes, to choose tightly or loosely defined classes (in a spectral variance sense), as well as to include or exclude mixtures of materials in samples. For example, given that one of the class categories of interest is grass, do we define a number of tightly defined grass prototype classes with a small variance (that we will later on consolidate into a single grass category after the classifier is finished) or do we combine all the grass samples into one grass prototype class that will exhibit a larger (perhaps very large) variance? As another example, given that the class of concern is swamp, do we define a number of swamp prototype classes (representing various mixture ratios of water and vegetation) or do we exclude this category and later on apply a mixed-pixel algorithm to the rejected pixels?

Optimal selection of class prototypes would seem critical to achieving optimal results from a supervised classifier. However, from an operational point of view, a key concern is whether it is possible for an analyst to identify the prototype classes needed in a timely manner, without too much difficulty, and without requiring an unusual amount of skill. Therefore, it is important to simulate varying degrees of operator skill and/or effort, investigating the consistency of performance results.

In most situations, an analyst will likely find it difficult to define all at once a complete set of prototype classes that is truly representative of a scene. There are two primary reasons for this difficulty. The first reason is that the analyst is unlikely (except in the case of very simple scenes) to be aware of all the natural and manmade features that exist within the scene, and even if the analyst was aware, a complete set of good samples are often difficult to find. The second reason is that a scene will seldom be a clean display of perfectly homogeneous and spectrally well-separated materials. Certain natural and manmade features are mixtures of materials.

This predicament strongly suggests the need for an iterative methodology. As the classifier processes data within a scene and encounters pixels that do not correspond to one of the prototype classes, it should have the ability to reject them. Rejected pixels could be subsequently processed in a number of alternative ways. In a most simple manner, the rejected pixels could be processed in another pass; whereby, new classes are added to the prior set of prototypes classes and such a new set of class prototypes used as the training model. Alternatively, the rejected pixels (now representing a relatively small portion of the original scene) could be clustered. More sophisticated processing could consider the rejected pixels as candidates for mixtures of the class prototypes.

As part of the optimal selection process, outlier pixels should be removed from training samples (if they are present) before the covariance matrices are computed and input to the training model. Outliers can occur, for example, when an operator mistakenly crops the boundary of a training area to include part of another feature, or perhaps a few scattered single pixels are located within an otherwise homogeneous area. The presence of only one to three outliers can seriously degrade the

estimate of the covariance parameters of the model. This issue is discussed further in the next section.

Another issue similar to outliers is the situation where a training set actually consists of two or three spectrally well-defined materials. Perhaps it is impossible for an analyst to physically draw a boundary between such materials of interest because the pixels are intermixed. If the analyst knows the area consists of a certain (small) number of materials, a simple clustering algorithm (such as KMEANS) should be able to sort the pixels and form the appropriate number of homogeneous training areas.

2.2.2 Improving Performance of Conventional Algorithms

On a number of occasions prior to and during this effort, the investigators have experienced performance problems with the Bayesian and Mahalanobis classifiers with regard to certain features. For example, these classifiers almost always have a higher error rate for water than does the far less sophisticated Euclidean minimum distance classifier. Also, at times the LAS software used at TEC generates non-fatal (but alarming) error messages regarding the possible singularity of some class covariance matrices.

The problem is addressed by attributing this difficulty to degenerate covariance matrices, resulting from insufficient variation in a sample (attributable to lack of feature variation and/or quantization effects), and proposing that all class covariance matrices be forced to have a certain minimum variance. In particular, it can be observed that water classes often have variances less than one. With such a small variance, the covariance factor in the classifier's discriminant function causes the algorithm to form a sort of impenetrable barrier that causes many legitimate water samples that are only a distance of 2-3 gray shade values from the components of the water class mean vector to be assigned to some other class that may actually be a distance of 20-40 gray shade values per component.

Improvements to the performance of the quadratic classifiers can also be made by removing outlier pixels from training samples (if they are present) before the covariance matrices are computed and input to the training model. Although the estimates of mean vectors are not significantly affected by a few outliers, the presence of outliers in a training sample can seriously corrupt the covariance estimates. Samples with only a very few outliers, say 2 to 3 percent, will grossly overestimate the underlying parent populations; particularly, if the outlier samples are from a material with a spectral signature quite different from the material of interest. For example, using Landsat Thematic Mapper data, 3 pixels of vegetation embedded in a sample of 100 water pixels would sharply increase the estimates of the population covariance matrix elements involving bands B4 and B5 (σ_{44}, σ_{45}, σ_{55}, etc.). This outlier effect is easy to show, for example, by using a microcomputer spreadsheet program and computing the variances for a sample of about 100 pixels, with and without a couple of outliers. The removal of obvious outliers, if they comprise a small percentage of the training data, should be simple to automate.

2.2.3 Handling Mixtures of Materials

The most challenging problem is to find a mechanism for recognizing the existence of mixtures, and identifying the elements and corresponding proportions within these mixtures. Given that a scene consists of pure pixels of materials and the caveats mentioned above in Sections 2.2.1 and 2.2.2, most conventional algorithms, including the simplest, will perform rather well. However, once mixtures of materials (impure pixels) are introduced, the difficulty of the problem increases many fold.

[illegible]

[illegible]

[illegible]	[illegible]
[illegible]	[illegible]
[illegible]	[illegible]
[illegible]	[illegible]
[illegible]	[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]	[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

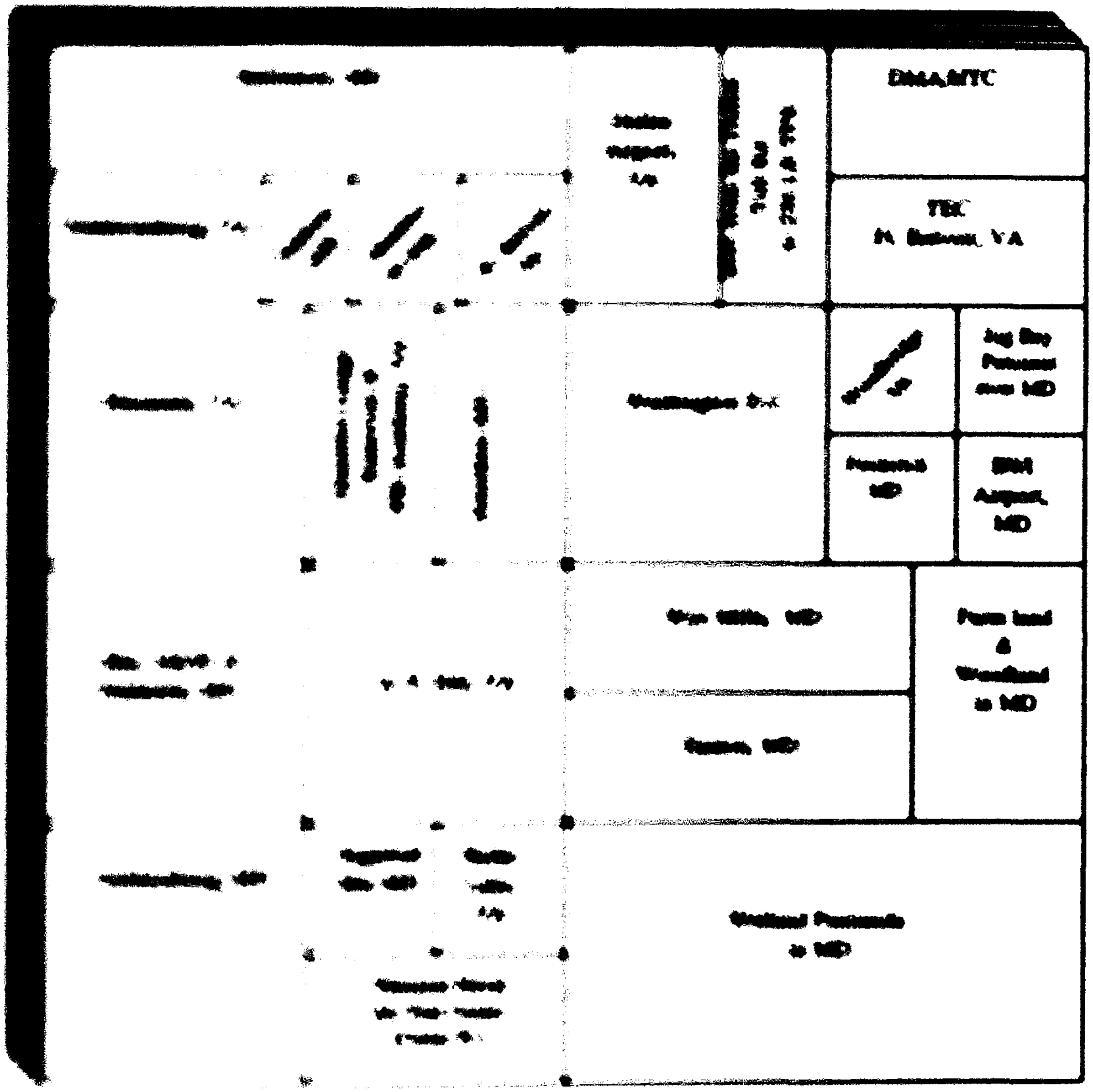


Figure [illegible]. Multidate/Multiframe Montage TM Data Set

During the course of this effort, all five dates of Landsat TM imagery were used. Initial trials focused on the May 1987 image. Once the behavior of the algorithms for this single date was established, the investigation proceeded to the remaining four dates.

Trials were conducted using a combination of training, test, and ground truth data extracted from the montage image set. During some trials, the actual montage image was classified and numerical accuracy assessed by comparing to a ground truth mask using the LAS system. During other trials, numerical accuracy was accessed by classifying the training data (autoclassification), test data, and ground truth data, which were extracted from the montage images using TEC-developed software on a microcomputer. Any data labeled as ground truth was verified by a personal site visit to the area.

Perhaps the easiest way to understand how this combination of data was used is to consider that all these data (training, test, and ground truth) were derived from a single large pool of data, into which the investigators placed their specific datasets. At various times during the effort, investigators extracted samples from the montage images with some knowledge of each site known through personal experience, analysis of the high resolution aerial photographs, map information, or personal site visit. Rather than give a historical chronicle of the training, test, and ground truth site extractions and of how the experiments were performed, we organized the description and results of the experiment by theme.

Some of the samples represent sites extracted with a high degree of skill or knowledge (sometimes with collateral high resolution photography), whereas others represent sites extracted with less skill or knowledge. Any of these sites would be valid candidates for training data and allow the testing of algorithms on highly skilled versus less-skilled site selection. The sites collected with a high degree of knowledge/skill would be valid for training or test data, whereas ground truth data (although sometime located by aerial photographs) were verified by site visit.

3.2 Training, Test, and Ground Truth Selection

As just discussed, the training and test data were extracted from a large pool of data that can be grouped into numerous candidate classes/sites. Each site (over 300 available in this pool) corresponds to a geographic site. The largest number of sites are defined by a LAS statistics file called MOSAIC.STATS that contains a collection of 296 sites. The sites were extracted, later examined by graphical and statistical analysis, and categorized into a smaller number of classes. Various descendents of the MOSAIC.STATS file were generated, resulting in statistics files with as many as 99 classes and as few as 10 classes. These files, along with a few other class/sites defined by another investigator in another file, comprise the pool of source data from which training and test sites are extracted and defined.

No sane person would attempt to use this particular method of site selection in a production environment. However, for the purpose of this study where we attempt a general characterization of the algorithms and test for robustness, this approach is really essential. Some scatter diagrams and graphs of spectral signatures are shown in Section 4.1 (Figures 3 to 11). In addition to portraying the layout of certain prototype classes in spectral space and indicating their separability, these figures also raise the concern of whether to include a small or large number of training sites and would seem to suggest that a rigorous analysis of a large set of prototypes is warranted. However, keep in mind that the ultimate intention is to define the simplest method for extracting training sites without compromising the classifier's accuracy.

As mentioned before, an attempt is made at distinguishing performance [illegible] defined by varying degrees of rigor. A numerical scheme is used to trace the [illegible] Classes 1 to 13 were selected quickly, based simply on knowledge of the area and the [illegible] appearance of the site in the multispectral scene. They are not part of the MOSAIC [illegible] was rigorously analyzed. Of these, classes 1 to 8 are spectrally homogeneous and are training classes that represent materials as opposed to cartographic features. Classes 100 to 200 are [illegible] that reside in the large class statistics file MOSAIC.STATS. Classes greater than 200 are ground truth sites verified by high-resolution photography and site visit.

The pool of data was used to construct four data sets called Dataset A, Dataset B, Dataset C, and Dataset GT. During the course of the experiments, Dataset A was used as a training dataset. Dataset B and Dataset C were used either as training data or test data depending on the [illegible] Dataset GT was defined as ground truth and used exclusively as test data. Except for the brief discussion below, the use of various combinations of these datasets will be discussed in Section 3.4.

Dataset A consists of nine classes that were given three different permutations during the course of the experiments. These permutations are given the names Dataset A1, Dataset A2, and Dataset A3 and are listed in Table 3-1. As mentioned, these datasets were used exclusively as training data. The purpose of this dataset is to test the performance of the classifier when the number of classes is kept to a minimum, and the selection is made to represent spectrally homogeneous classes that represent materials (rather than cartographic features such as roads, urban, marsh, agricultural fields, etc). The working hypothesis is that the objects within a scene (e.g. cartographic features) are actually composed of a small variety of spectrally-unique materials and that the large amount of spectral variation is due to mixtures of materials. Although at finer spectral resolution there is perhaps a large variation of fine spectral detail within the various materials, it is [illegible] level of classification needed, these variations can be ignored.

Dataset B consists of 26 classes that were given two permutations during the course of the experiments. These permutations are given the names Dataset B1 and Dataset B2, and are listed in Tables 3-2 and 3-3. Dataset B1 contains 20 classes and was used as a training set. Dataset B2 contains these same 20 classes (Classes 100-194) plus an additional six classes ([illegible]); however, note the data from these classes were sampled so that no class contained more than [illegible] samples. A somewhat different working hypothesis (from that in Dataset A) was used in that these classes correspond to cartographic features that may or may not be pure materials. Classes from this dataset were sometimes used for training and sometimes used as a test dataset. Some of these classes were also used to study the statistical properties of some of the class distributions.

Dataset C contains 25 classes and was used as a source for some of the graphical data and the mixture analysis. The original intention was to use these classes as another source of training and test data for further classification runs; however, the study was becoming extensive and it was decided to halt the classification trials in favor of performing the mixture analysis. A description of these classes is listed in Table 3-4. For the most part, these classes are individual (or a small number of) geographic sites extracted from within the broader classes in Dataset B.

Dataset GT contains eight classes and was used as test data for some of the trials. A description is given in Table 3-5.

Appendix A provides supporting statistical data for the trials. In this appendix, Tables [illegible] list the mean vectors for the classes in Datasets A and B; Table A7 lists the covariance matrices for the classes in Dataset A; and Table A8 lists the correlation matrices for the classes in Dataset [illegible]

Table 3-1 Classes in Datasets A1, A2, A3

Dataset A1 (Training)

Class	Name	Description	# of Samples
1	Water 1	Light Blue Water	100
2	B. Roof	Bright Metal Rooftop	26
3	D. Veg	Deciduous/Bright Red Vegetation	60
4	C. Veg	Coniferous Vegetation	60
5	Asphalt	Dulles Airport Parking Lot	70
6	Concrete	Concrete from Andrews AFB	81
7	Water 2	Dark Blue Water	60

Dataset A2 (Training)

Class	Name	Description	# of Samples
1	Water 1	Light Blue Water	[illegible]
2	B. Roof	Bright Metal Rooftop	[illegible]
3	D. Veg	Deciduous/Bright Red Vegetation	[illegible]
4	C. Veg	Coniferous Vegetation	[illegible]
5	Asphalt	Dulles Airport Parking Lot	[illegible]
6	Concrete	Concrete from Andrews AFB	[illegible]
7	Water 2	Dark Blue Water	[illegible]
120	Grass-A	National Airport Grass	[illegible]

Dataset A3 (Training)

Class	Name	Description	# of Samples
1	Water 1	Light Blue Water	[illegible]
2	B. Roof	Bright Metal Rooftop	[illegible]
3	D. Veg	Deciduous/Bright Red Vegetation	[illegible]
4	C. Veg	Coniferous Vegetation	[illegible]
5	Asphalt	Dulles Airport Parking Lot	[illegible]
6	Concrete	Concrete from Andrews AFB	[illegible]
7	Water 2	Dark Blue Water	[illegible]
122	Grass-B	National Bureau of Stds. Grass	[illegible]

[illegible]

[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]		[illegible]
[illegible]	[illegible]	[illegible]		[illegible]
[illegible]	[illegible]	[illegible]		[illegible]
[illegible]	[illegible]	[illegible]		
[illegible]	[illegible]	[illegible]		[illegible]
[illegible]	[illegible]	[illegible]		
[illegible]	[illegible]	[illegible]		[illegible]
[illegible]	[illegible]	[illegible]		[illegible]
[illegible]	[illegible]	[illegible]		[illegible]
[illegible]	[illegible]	[illegible]		
[illegible]	[illegible]	[illegible]		
[illegible]	[illegible]	[illegible]		
[illegible]	[illegible]	[illegible]		[illegible]
[illegible]	[illegible]	[illegible]		[illegible]
[illegible]	[illegible]	[illegible]		

[illegible]

[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	

[illegible]

[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]		[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

[illegible]

3.5 Linear Mixing Trials

The linear mixture analysis was directed at swamps, which can presumably be modeled as mixtures of vegetation and water. The trials address two questions:

> (1) Is it possible that endmembers other than water and vegetation can be used to adequately model swamp?
>
> (2) Is it possible to distinguish the type of vegetation (e.g. grass, deciduous trees, coniferous trees) that is present in the mixture?

Closely related to these questions is the issue of nonunique solutions, which is explored in detail.

Ten samples were selected to test the linear mixing model. These were extracted from Dataset B2, and Dataset C, and are identified as follows:

Label	Material
C174	Swamp
C175	Swamp
C176	Swamp
C123	Grass
C125	Grass
C133	Leaf
B140	Pine
B160	Asphalt RW
B162	Concrete
B190	Water

The swamps, C174, C175, C176, are the materials assumed to be mixtures of water and some type of vegetation. The remaining materials are tested as possible endmembers.

The analysis focused on an approach that begins with pairwise combinations of candidate endmembers, and expands the model to include additional endmembers only if the best pairwise model is inadequate. Prior to this, trials that considered full regression model combinations of three to four endmembers were tested, and a standard method of model reduction was attempted. This alternative approach seemed to offer no advantage over the approach that begins with pairwise endmember combinations, and had a number of disadvantages, including too few degrees of freedom for the residual sum of squares, the possibility of negative coefficients (implying a negative amount of the corresponding material), and problems of imposing the physical constraints mentioned in Section 2.1.5.

The trials began with determining the domain limits defined by each of the pairs of endmembers. These limits must necessarily be considered approximate because sample mean vectors for each of the endmembers were used in the definition, and since each sample is a cloud of data, there are obviously individual endmembers in each sample that would increase the width of the domain/interval. A better method of defining the interval would perhaps be to choose the extremums of the data cloud, so long as these extremums were not outliers. However, this would have increased the complexity of implementing the trials beyond what could be allocated to the current effort. Such a method should be tested in the future.

The domain/interval limits were used to assign a degree of compliance (DOC) with the first physical constraint to restrict the allowable endmember combinations. Regression models are then computed with diagnostic statistics for each of the pairwise endmember combinations. An F-ratio is used to assess the statistical significance of a model. If none of the candidate endmember pairs had produced a statistically significant model, then the model would have been expanded to include additional endmembers (up to a 4-component model).

The selection process employed four criteria: (1) suitable endmember combinations need to have a high DOC with the first constraint; (2) large F-ratio models were considered superior to smaller ones in a statistical sense; (3) the model needed to be physically relevant by passing the second constraint that all model coefficients were positive and sum to approximately the value of one, as mentioned in Section 2.1.5; (4) each and every residual must be small.

Results are discussed in Section 4.7.

4.0 DISCUSSION OF RESULTS

4.1 Graphical Analysis of Real-World Spectral Signatures

Before delving into the computational analysis that was performed, lets attempt to gain insight into the spectral nature of the features being studied by visually examining some graphical presentations of the data. Just as a picture can be worth a thousand words, so can it be worth just about that many numbers.

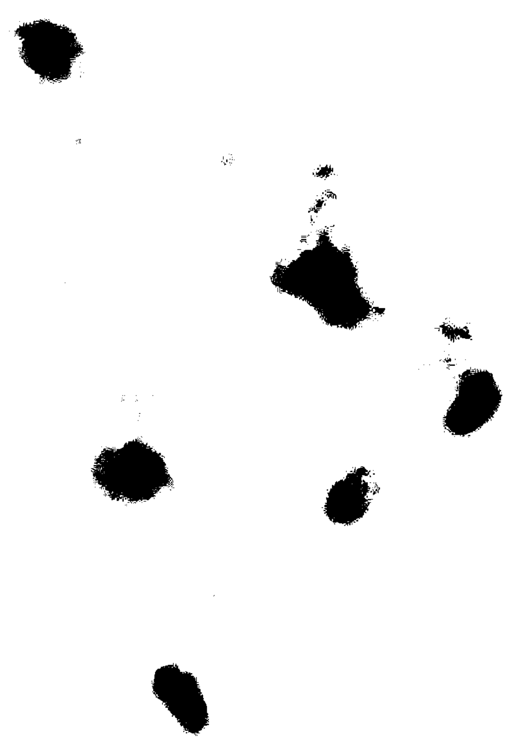
The data are presented in two ways. Figures 3 and 4 are projections of three-dimensional scatterplots of data derived from some of the training classes that were used to test the classifier's performance. Figures 5 to 12 are graphs of signatures derived from a few representative training and ground truth sites.

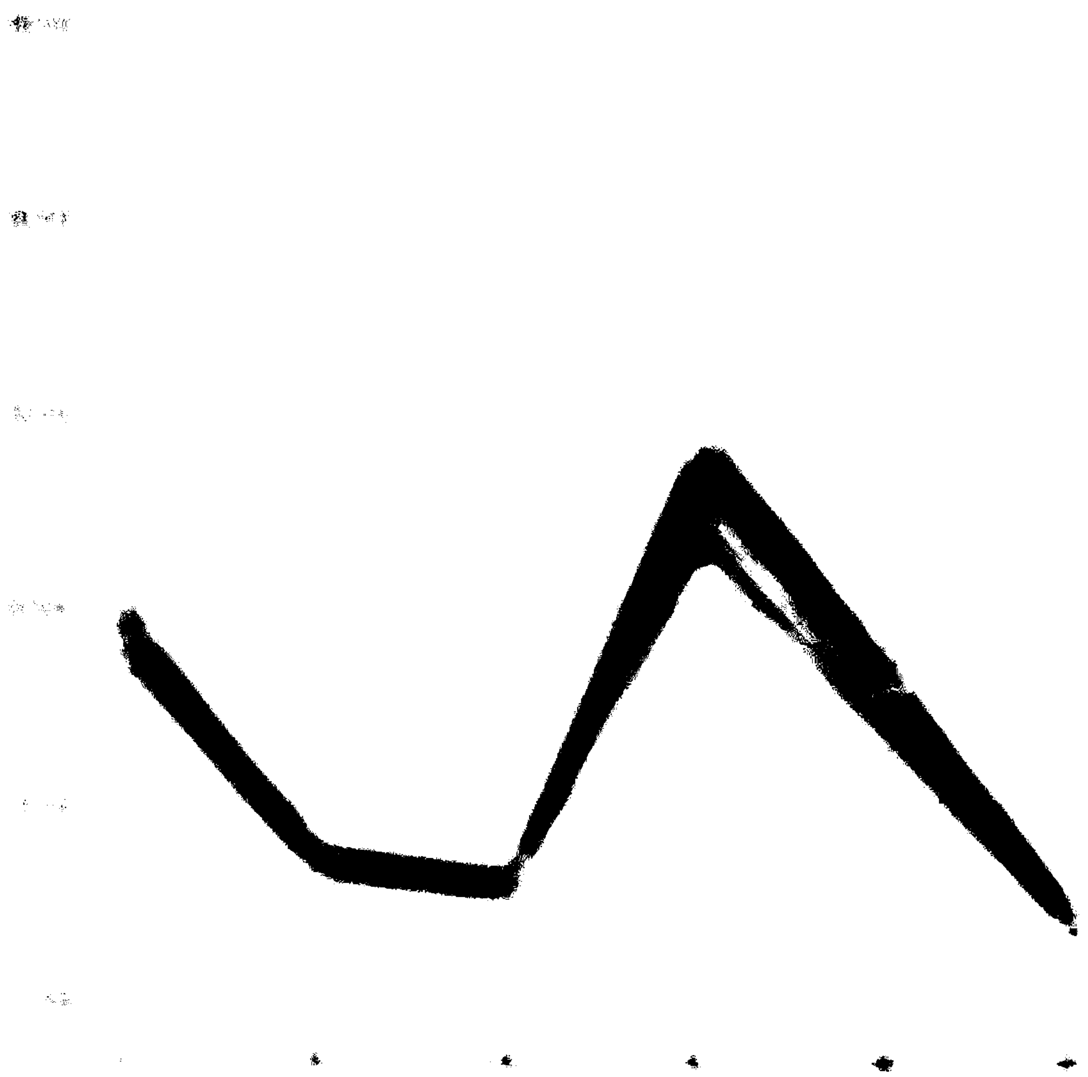
Observing the scatterplot projections in Figures 3 and 4, one property that becomes immediately obvious is how samples from concrete, asphalt, water, deciduous trees and coniferous trees are easily separable in spectral space. The samples from each of these classes form well-defined clusters that do not overlap.

Notice the two separate clusters for the classes Grass-A and Grass-B. The first thing to notice is that even though both classes are grass, they occupy a different portion of the spectral space. If these two classes were combined into a single training class, the resulting pooled covariance would be quite large and likely lead to confusion with the deciduous trees class. Therefore, the graphical analysis indicates that they should not be combined.

The second thing to notice about classes Grass-A and Grass-B is that if a line is drawn between the Concrete and D. Veg centroids, the two grass classes lie on this line. This is true for either Figure 3 or Figure 4, each representing different projections in spectral space. Most notable is the observation that Grass-A appears to be located midway on the line connecting Concrete and D Veg. Since concrete spectra often resembles soil spectra, this grass sample probably has a significant soil component; i.e., it is a mixture of vegetation and soil. Therefore, two interpretations can be given to Grass-A. The first is that this class represents a single material (grass) with its own rightful place in spectral space, whereas, the second is that this class is a mixture of two *endmember* classes, pure grass and pure soil. The second scenario can easily occur for unhealthy or dying grass with a relatively low biomass (compared with healthy well-maintained grass) where a good amount of soil reflectance is present. It is worth mentioning at this time that the use of Grass-A as a training class in Trial 2 resulted in poor performance. In particular, numerous test samples within the TEC, High School, and Mall sites (that should have been labelled concrete) were misclassified as Grass-A.

Notice that if a line is drawn between the centroids of D. Veg and Water 1 in either Figure 3 or Figure 4, the samples of C. Veg lie very close to this line and that they are also almost midway between D. Veg and Water 1. In this case, it can be assumed that C. Veg corresponds to a particular form of vegetation (coniferous) and that it is not a mix of deciduous vegetation and water. However, suppose we introduce a swamp class that is indeed a mixture of D. Veg and Water 1. It is very conceivable that this class will occupy the same portion of spectral space. This apparent overlap will also be later confirmed when Figures 5 to 8 are examined. In fact, this phenomena offers an explanation for the confusion observed to occur between swamp and coniferous trees in the classification trials. The graphical nature of the data is strongly suggesting a possible degeneracy in the spectral space defined by this small number of bands.





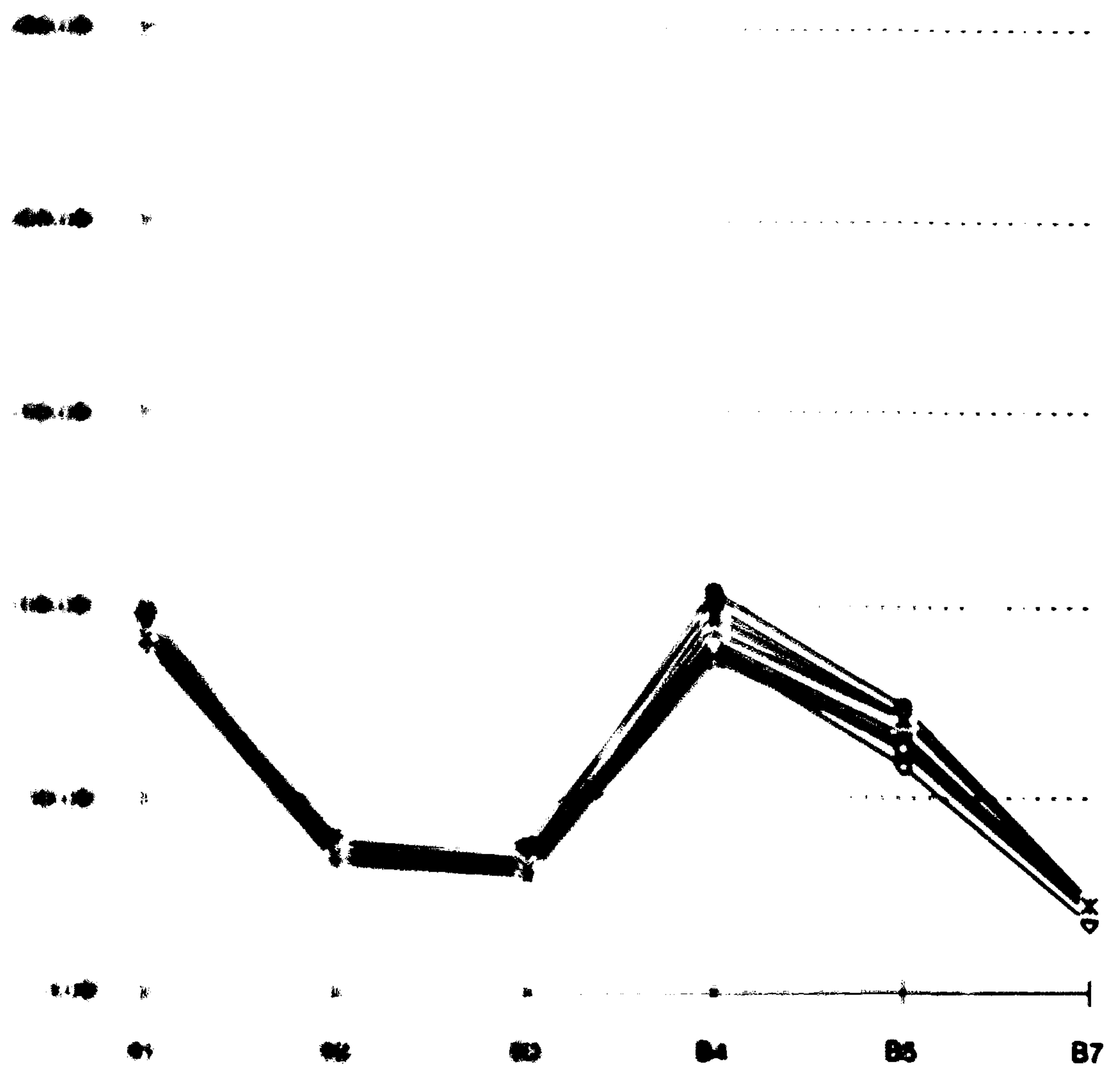


Figure 6. Spectral Signatures of Coniferous Trees

This scattering component is also present in the signatures of Figures 6 to 12. Fortunately, because the scattering component is additive, it will not affect the separability of the training classes or the performance of the classifiers (unless the scattering is nonuniform in the scene, which was only true for the August 1987 mosaic scene).

Figure 6 shows the mean spectral curves for 10 sites of coniferous trees. As was the case for deciduous trees, the general trend (shape) for all these sites is similar. The greatest intensity response and variation occurs once again in B4, due to the reflectance properties of chlorophyll; however, it is somewhat less in intensity than for deciduous trees. The intensity and variation in B5 is about the same as that for deciduous trees.

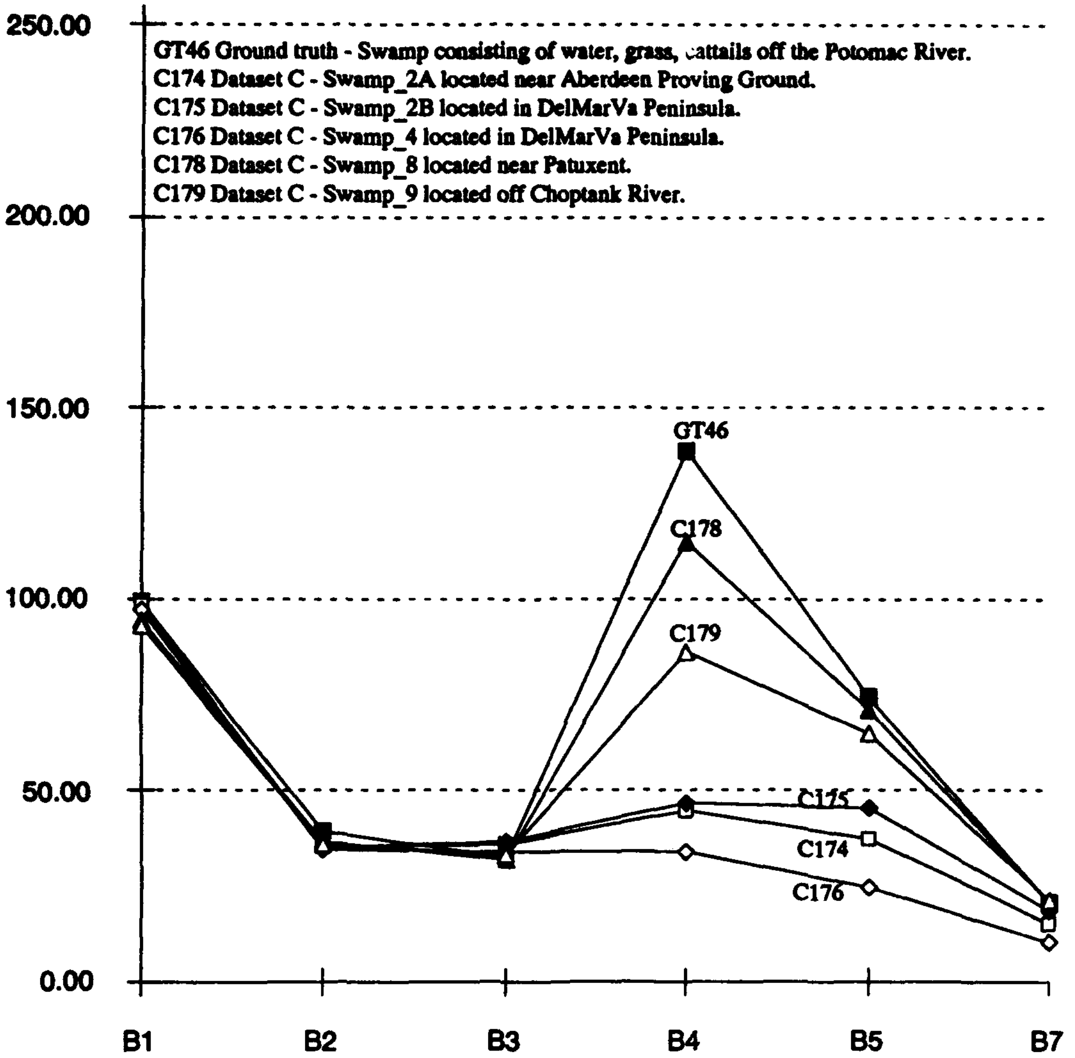


Figure 7. Spectral Signatures of Swamp Sites MY85

Figure 7 shows the mean spectral curves for six MY85 swamp sites. Unlike the previous graphs for deciduous and pine sites, the curves of these sites do not follow the same trend. This is particularly true for the spectral region represented by bands B3 to B5. Not only is there a large variation in the intensity variations of bands B4 and B5, but there are significant variations in the slopes of the curves between B3 to B5.

These variations are indicative of different mixing proportions in water and vegetation (along with perhaps different species of vegetation) that compose the swamp sites. Although Swamps C174, C175 and C176 occupy a separate region of spectral space from the other classes considered, others do not. Note the overlap between the GT46 swamp and deciduous trees (Figure 5), the C178 swamp and deciduous trees (Figure 5), and the overlap between the C179 swamp and coniferous trees (Figure 6).

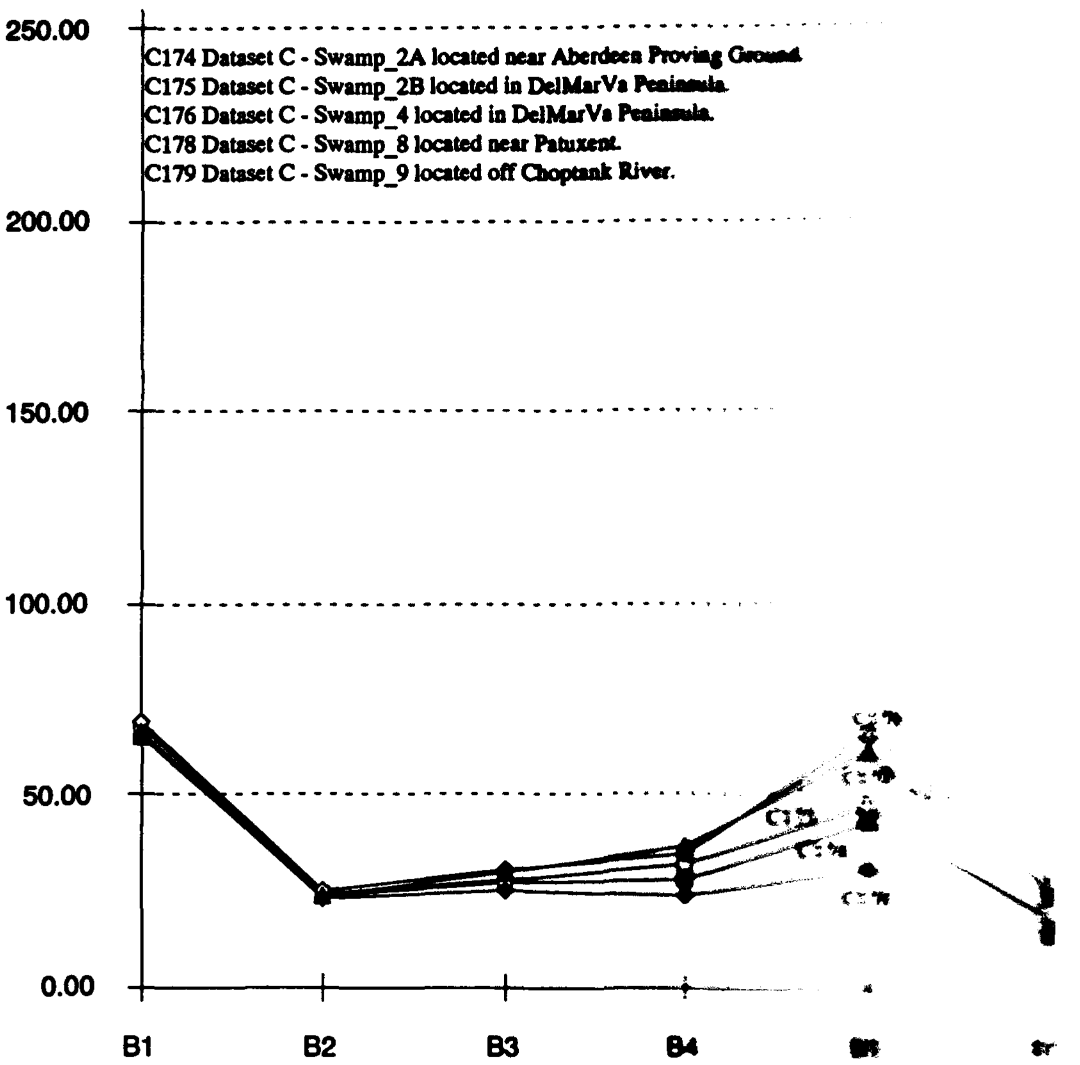


Figure 8. Spectral Signatures of Swamp [illegible]

Figure 8 shows the mean spectral curves in October 1985 for five [illegible] not available) displayed in Figure 7. In addition to showing [illegible] season, the responses in October (particularly in B4) can be used to [illegible] different ground feature from coniferous trees, and C178 is a [illegible] deciduous trees. For example, observe the following differences.

<u>MY85</u>	<u>B1</u>	<u>B2</u>	<u>B3</u>	<u>B4</u>	<u>B5</u>	[illegible]
Swamp-C179	92.88	35.81	33.59	[illegible]	[illegible]	[illegible]
PINE	93.92	35.14	31.65	[illegible]	[illegible]	[illegible]

<u>OC85</u>	<u>B1</u>	<u>B2</u>	<u>B3</u>	<u>B4</u>	<u>B5</u>	[illegible]
Swamp-C179	66.46	24.13	29.81	[illegible]	[illegible]	[illegible]
PINE	61.46	22.06	20.53	[illegible]	[illegible]	[illegible]

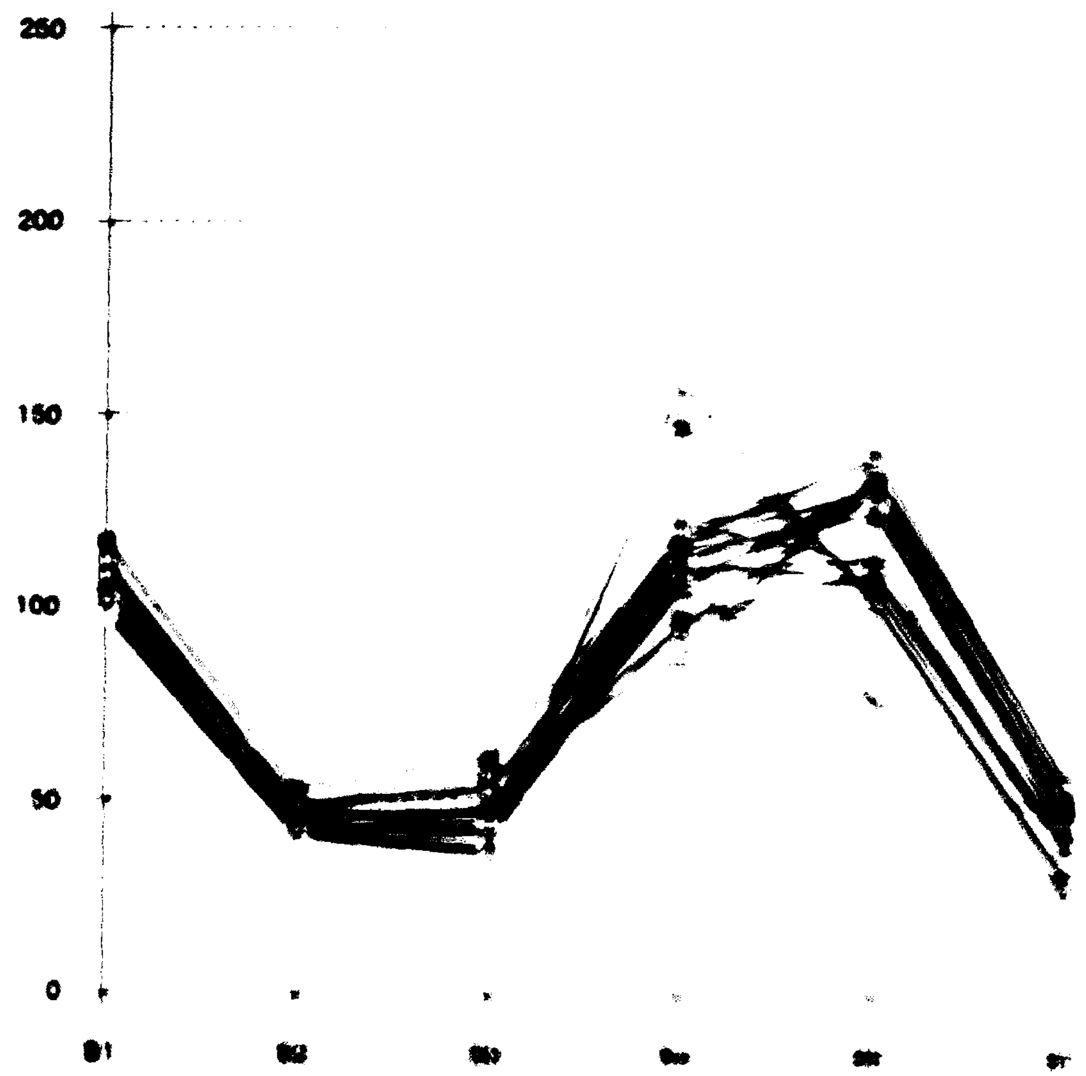
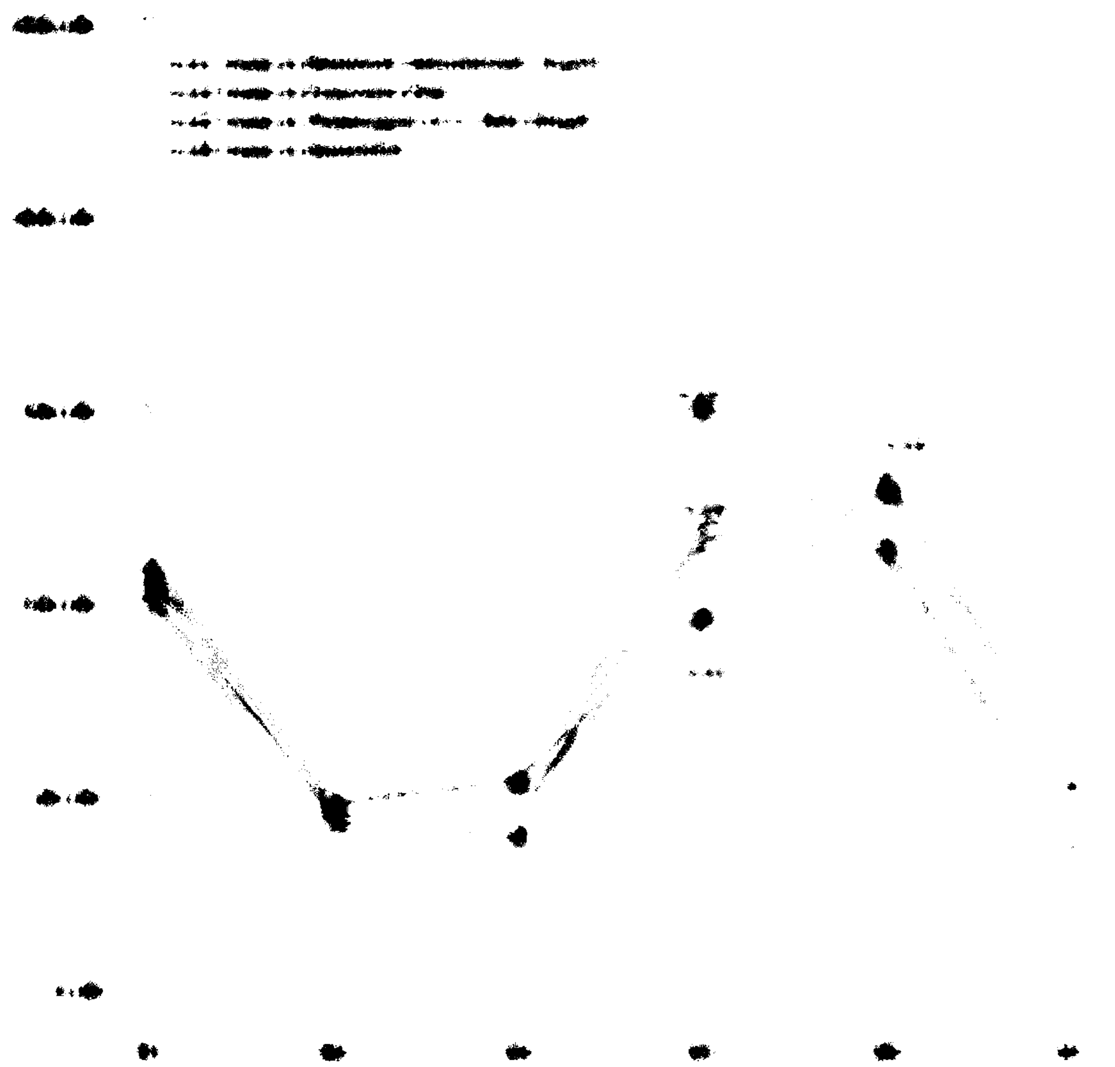
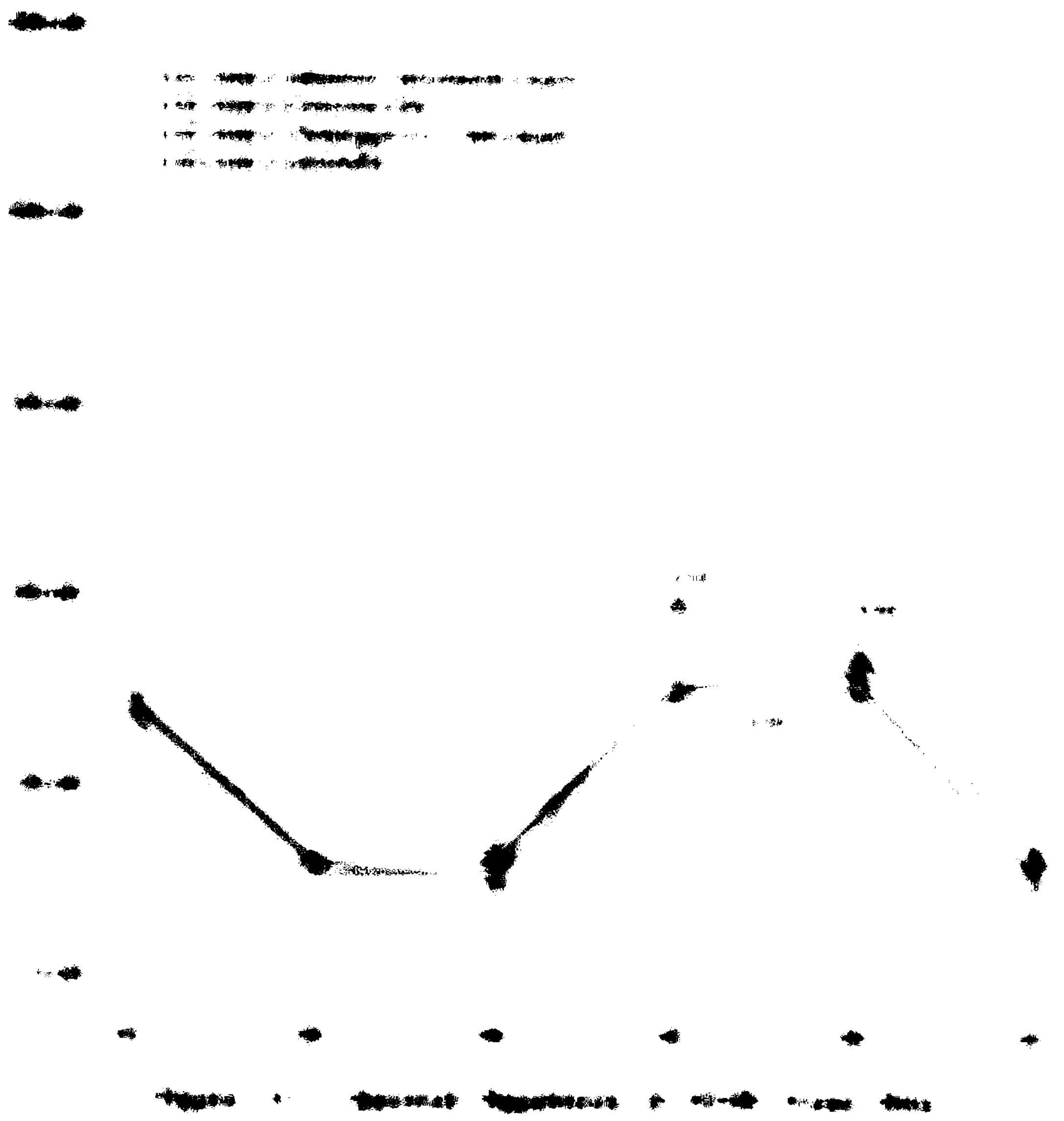
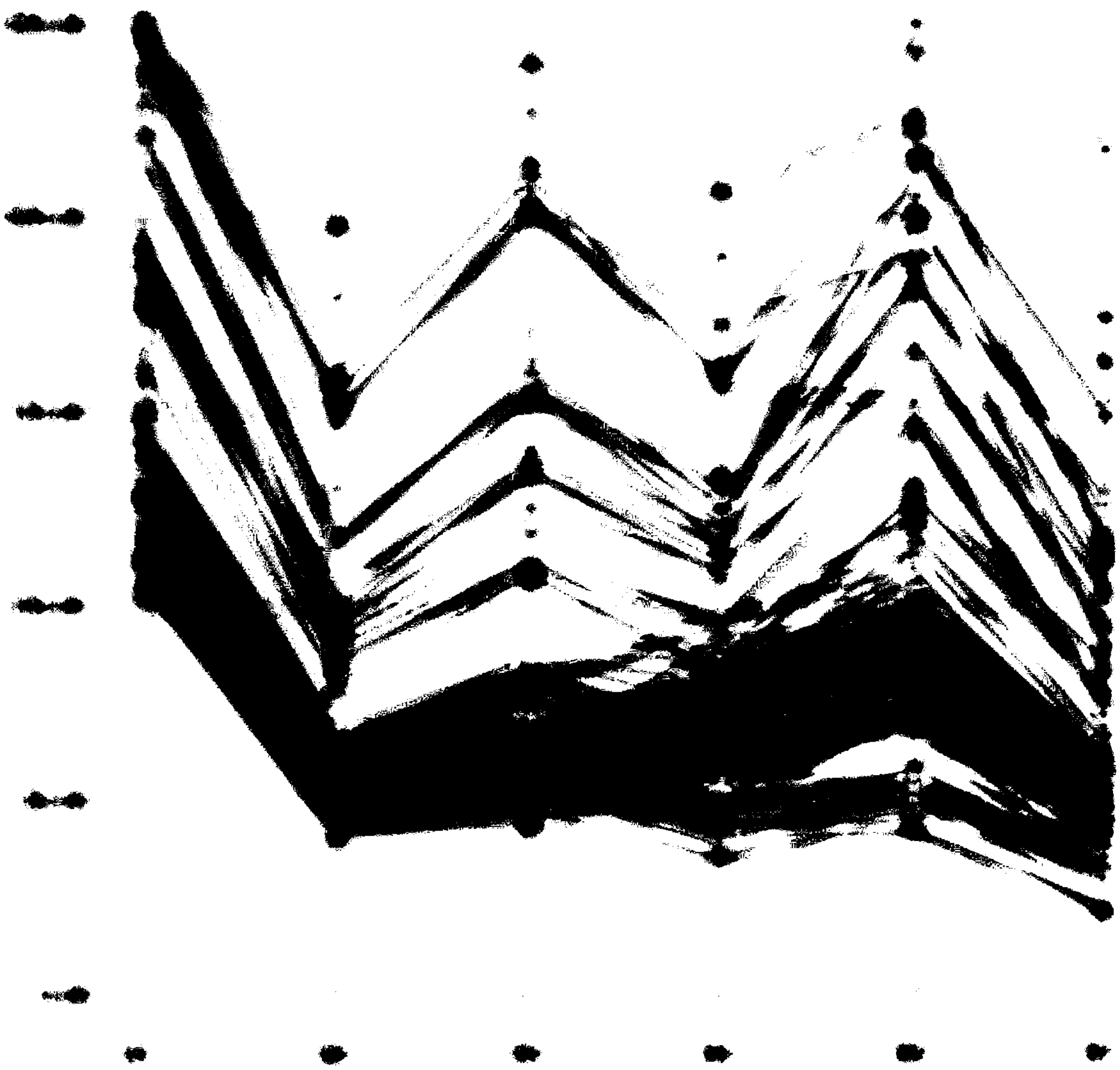


Figure 9. Spectral Signatures of [illegible] Grass Sites

Figure 9 shows the mean spectral curves for [illegible] grass sites. The variation of grass includes healthy, well-maintained grasses, sod farms, less healthy grasses, and pastures. Notice the high variation of responses (particularly in B4 and B5) and degree [illegible] for the different grasses. All these sites were visited in person and verified as grass. Unfortunately, because of the elapsed time between the scene's acquisition and the site visit, as well as the tendency for grass to change dramatically in short periods of time due to variations in weather and maintenance, it is not possible to identify a precise cause and effect relationship for the spectral variations. However, a reasonable explanation is to attribute the variation to varying amounts of vigor and/or biomass for the grass, as well as to the amount of thatch and soil present in the samples.







The possible degeneracy in the spectral space, where a mixture of materials combines to form a signature identical to certain pure pixels, may pose a serious problem for distinguishing some [illegible] and [illegible] features. The graphical analysis has revealed specific cases where this degeneracy can manifest. When such situations exist, no algorithm, regardless of its complexity will separate them. The spectral information just simply doesn't exist to distinguish them. Solving the problem would require increased spectral or spatial information.

The addition of more spectral bands with increased spectral resolution, hopefully, can eliminate the degenerate cases. However, there is no guarantee that this approach will be successful. The underlying spectra might be quite bland and not contain distinguishing absorption features. Therefore, incorporating such data, although more voluminous, would not necessarily provide increased spectral information.

In such cases, values for the coefficients in the vector $\hat{\beta}$ of the linear mixing model could conceivably be found that generate a mixture spectra almost identical to some other mixture spectra, or to some other pure pixel. Assuming there is almost an infinite number of candidate endmembers in the real world that can combine linearly to produce a mixture spectra, there is almost an infinite number of candidate $\hat{\beta}$ vectors, any one of which could produce identical spectra, and, therefore, a degenerate spectral space.

4.2 Methods to Assess Classification Accuracy

The results of the classification runs were initially assembled into contingency tables that show the results in detail (see Appendix B). Each row of the table corresponds to a test class, and the columns list the number of samples placed into each of the prototype classes.

The contingency table results are summarized by tables in this section, which list omission and commission errors. Each type of error takes a different view of the results. **Omission error** is from the viewpoint of the test (ground truth) data. Given a group of test (ground truth) data, how many samples did the classifier mislabel as something else? For example, if there are 100 water samples in the test data and 5 of the samples were misclassified, the omission error would be 5 percent. **Commission error** is from the viewpoint of the resulting class map. Given that the classifier labeled a certain number of samples as a particular category, how many of these samples correspond to something else? This error gives the false alarm rate. For example, if the classifier labeled 100 samples as water and 2 of the samples were actually something else (according to the test data or ground truth), the commission error and the false alarm rate for this category would be 2 percent.

Although the groupings of test data remain a constant for all the various classification trials, the groupings of the class map data are not constant. Therefore, comparing omission error results as percentages is a reasonable thing to do; however, comparing commission error results as percentages can be misleading. In comparing two trials, the percentage of commission errors could conceivably increase, even though the absolute number of commission errors decreases dramatically. This is discussed further in Section 4.3, where this situation occurs during Trial 3.

In comparing the class names for training sites with those of the test site, one quickly notices that there is not always a one-to-one correspondence. For example, the test class *Mall* does not correspond to any of the training classes in Datasets A1-A3. However, for our purpose, we could consider the classifier to be correct if it labeled such pixels as either asphalt or concrete since it is quite conceivable that a shopping mall would be an aggregate of asphalt and concrete materials.

In order to conduct a quantitative analysis, some kind of equivalence musted be established between the classes in the training sets and those in the test sets. Of course, in the case of auto-classification, such a correspondence is automatic, and in some test classes the correspondence is immediately obvious.

Tables 4-1 and 4-2 define the equivalence between training and test classes that are used to summarize the omission and commission results as presented in the following section. The omission and commission results are computed from the contingency tables listed in Appendix B (Refer to this appendix for a detailed look at the classification results).

Table 4-1 Class Equivalence Sets for Omission Errors for Trials 1-4

Construction =	{Asphalt}
TEC Site =	{Asphalt, Concrete}
Parkland 1 =	{D. Veg}
High School =	{Asphalt, Concrete}
Mall =	{Asphalt, Concrete}
Parkland 2 =	{Grass-A, Grass-B}
Baresoil =	{Concrete}
Fields-A =	{Grass-A, Grass-B}
Fields-C =	{Grass-A, Grass-B}
Fields-D =	{Grass-A, Grass-B}
Grass-A	{Grass-A, Grass-B}
Grass-B =	{Grass-A, Grass-B}
Grass-C =	{Grass-A, Grass-B}
Leaf =	{D. Veg}
Pine =	{C. Veg}
Road-A =	{Asphalt}
Runway-C =	{Asphalt}
Runway-F =	{Concrete}
Swamp-A =	{Water 1, D. Veg, C. Veg, Water 2, Grass-A, Grass-B}
Swamp-B =	{Water 1, D. Veg, C. Veg, Water 2, Grass-A, Grass-B}
Urban-D =	{B. Roof, Asphalt, Concrete}
Urban-F =	{B. Roof, Asphalt, Concrete}
Urban-I =	{B. Roof, Asphalt, Concrete}
Water-A1 =	{Water 1, Water 2}
Water-A2 =	{Water 1, Water 2}
Water-C =	{Water 1, Water 2}

Table 4-2 Class Equivalence Sets for Commission Errors for Trials 1-4

Water 1 =	{Water A1, Water A2, Water C, Swamp-A, Swamp-B}
B. Roof =	{ — }
D. Veg =	{Parkland 1, Leaf}
C. Veg =	{Pine}
Asphalt =	{Construction, TEC Site, High School, Mall, Road-A, Runway C, Urban-D, Urban F, Urban I}
Concrete =	{TEC Site, High School, Mall, BareSoil, Runway F, Urban-D, Urban F, Urban I}
Water 2 =	{Water A1, Water A2, Water C, Swamp-A, Swamp-B}
Grass-A =	{Parkland 2, Field-A, Fields-C, Fields-D, Grass-B, Grass-C}
Grass-B	{Parkland 2, Field-A, Fields-C, Fields-D, Grass-A, Grass-C}

4.3 Results of Trials 1 and 2

Trials 1 and 2 were preliminary trials conducted on a single scene (May 1987). The classifiers were applied to both the training and test data. These were simple runs intended to test the use of a small number of training classes. Trial 1 contains the 7 prototype classes in Dataset A1, whereas, Trial 2 contains the 8 prototype classes in Dataset A2. The distinguishing factor between these two trials is the addition of a grass class in Trial 2. The results are reported in terms of auto-classification errors and omission errors in Tables 4-3 and 4-4.

The auto-classification results for all three classifiers are excellent with 100 percent of all samples being labeled correctly. This indicates that the training classes are spectrally well separated. Consequently, the classifiers had no problem labeling its own training data correctly.

The performance degraded when the classifiers were applied to data outside the training data. According to Tables 4-3 and 4-4 the error rate remained low for some classes, however, it was quite high for certain other classes. In particular, note the high omission error for the Bayesian classifier of 76.4 percent and 66.7 percent for the Swamp-A and Swamp-B, respectively.

There is no corresponding swamp class in the training data, but recall from the class equivalence definition that the swamp data would have been considered correctly classified if it was labeled as *Water 1, Deciduous Vegetation, Coniferous Vegetation, or Water 2*. This equivalence is reasonable if one considers swamp to be a mixture of water and vegetation and that a subsequent algorithm would label such a mixture as swamp. However, the Bayesian and Mahalanobis distance classifier labeled the majority of this swamp data as asphalt. The Euclidean classifier labeled most of this class correctly.

Notice from the contingency tables B2 (i, ii, and iii), found in Appendix B, that the classifiers in Trial 1 usually labeled grass samples as D. Veg, which is a reasonable assignment given the alternatives without the grass class. Therefore, we can assume that if such an assignment is acceptable to the analyst (perhaps only to separate vegetation from urban and water), then it would not be necessary to train the classifier with a grass class.

The addition of a grass class in Trial 2 enabled the field and grass test data to be evaluated. However, a problem develops because the omission errors for the Construction, TEC Site, High School, and Mall increase. From the contingency tables B2 (iv, v, and vi), found in Appendix B, observe that a significant number of samples within these test classes are being labeled as grass.

It will be seen in Trial 3 that replacing the grass class with another grass class solves this problem. The implication is that one must be careful in selecting grass prototypes. Apparently, the grass class used in this trial had a soil, concrete, or asphalt component within it that made this class similar to asphalt or concrete. According to the auto-classification results, the Grass-A sample was still spectrally separable from the asphalt and concrete class prototypes; however, too many samples within the Construction, TEC site, High School, and Mall are closer to this Grass-A sample than Asphalt or Concrete.

[illegible]

[illegible]

[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]

[illegible]

[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]			
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]			
[illegible]			
[illegible]			
[illegible]			
[illegible]			
[illegible]			
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

With minor exceptions, the modified Bayes method improved the results of the standard Bayes classifier. The problem in omission errors for the water classes disappeared and the errors for the water-related classes were greatly reduced:

> Water-A1 improved from 16.90% error to 0.00% error.
> Water-C improved from 15.30% error to 0.00% error.
> Swamp-A improved from 66.47% error to 15.20% error.

Of course, the Swamp-A classes were not actually classified as swamp because there were no swamp prototype classes. They were classified as some type of water or vegetation (see the contingency tables in Appendix B and Tables 4-1 to 4-2 listing class equivalence sets).

This improvement corrected a major flaw of the standard Bayes algorithm. Reference the contingency results listed in Table B4 (iv and v) of Appendix B and notice that **a large number of the misclassified water and swamp samples were labeled as asphalt.** By invoking the minimum variance criterion all of the water samples were labeled correctly, and the number of swamp samples mislabeled as asphalt was reduced from 446 to 102.

The modified Bayes method also improved the commission results, or false alarms, corresponding to the asphalt class:

> Asphalt false alarms were reduced from 681 samples to 139 samples.

The false alarms for coniferous vegetation increased from 134 samples to 203 samples; however, this problem is not as bad as it appears. Referencing the contingency results in Appendix B, Table B4 (iv and v), notice that 180 out of these 203 samples belong to the test dataset's swamp class. This is a category for which there is no training class. Given that swamp can be defined as a mixture of vegetation and water and that thus far we have not invoked a rejection criterion, this assignment of swamp samples to coniferous vegetation can easily be considered correct. Of course, for subsequent trials where rejection criteria are tested, we should expect to see such false alarms disappear (this, in fact, does occur).

Numerous minimum variance threshold values were tested that ranged from 1.0 up to 25.0 (only a value of MinVar=16 for water and MinVar=3 on other classes is shown). The best results were achieved for the values shown. A larger value for water increased the errors for other classes, whereas a smaller value increased the errors for the swamp class.

The lack of swamp training classes was actually intentional for this trial. Other trials include this class. Consequently, the issue of whether to identify swamps using numerous training classes or using a mixture approach can be explored. Using the training class approach, many training classes for swamp are likely to be needed for a scene because of the large variations of possible mixtures (e.g. 80% water and 20% vegetation; 50% water and 50% vegetation; 20% water and 40% vegetation; etc.), not to mention the various possible species of vegetation.

If a mixture approach is attempted, one strategy would be to classify swamps as either water or vegetation, with the intention to reject by the chi-squared threshold. In rejecting the classification, but then remembering that the samples were rejected as a water or a vegetation classes, they could be tagged as such for mixed-pixel analysis. Subsequent analysis would then recognize the definition that swamp is a mixture of water and vegetation. However, if the samples were rejected, but remembered as asphalt, this strategy would fail.

Table 4-5 Auto-Classification Errors for Trial 3

This table lists the percentage of error in classifying the prototypes within each of the classes in the training set A3, using the Modified and Standard Bayes discriminant; the Mahalanobis distance; and the Euclidean distance methods.

PROTOTYPE	Modified Bayes	Standard Bayes	Mahalanobis	Euclidean
Water 1	0.00%	0.00%	0.00%	0.00%
B. Roof	0.00%	0.00%	0.00%	0.00%
D. Veg	0.00%	0.00%	0.00%	0.00%
C. Veg	0.00%	0.00%	0.00%	0.00%
Asphalt	0.00%	0.00%	0.00%	0.00%
Concrete	0.00%	0.00%	0.00%	0.00%
Water 2	0.00%	0.00%	0.00%	0.00%
Grass-B	0.00%	0.00%	0.00%	4.17%

Table 4-6 Commission Errors for Trial 3

This table lists the commission errors in classifying the test data test Set B2, using the Modified and Standard Bayes discriminant; the Mahalanobis distance; and the Euclidean distance methods. Training Set A3 was used to train the classifier. The modified Bayes was run using minVar =16 for the water classes and minVar =3 for all other classes. The commission errors were computed using the "class equivalence set for commission errors" listed in Table 4-2 and the contingency results listed in Table B4 of Appendix B. Both percentages and actual numbers of errors are given.

PROTOTYPE	Modified Bayes	Standard Bayes	Mahalanobis	Euclidean
Water 1	0.00%	0.00%	0.00%	0.00%
B. Roof	-----	----	----	----
D. Veg	22.74%	21.98%	21.33%	24.78%
C. Veg	34.64%	26.17%	23.73%	52.82%
Asphalt	19.83%	54.74%	55.67%	48.50%
Concrete	66.71%	67.00%	67.09%	45.43%
Water 2	0.00%	0.00%	0.00%	0.00%
Grass-B	29.48%	31.67%	34.10%	20.29%

PROTOTYPE	Modified Bayes	Standard Bayes	Mahalanobis	Euclidean
Water 1	0	0	0	0
B. Roof	-----	----	----	----
D. Veg	266	244	224	311
C. Veg	203	134	117	440
Asphalt	139	681	707	599
Concrete	523	530	532	159
Water 2	0	0	0	0
Grass-B	263	298	343	97

Table 4-7 Omission Errors for Trial 3

This table lists the omission errors in classifying the test data test Set B2, [illegible] discriminant; the Mahalanobis distance; and the Euclidean [illegible] classifier. The modified Bayes was run using minVar =16 for the [illegible] The omission errors were computed using the "class equivalence set for [illegible] the contingency results listed in Table B4 of Appendix B.

TEST SITE	Modified Bayes	Standard Bayes	Mahalanobis	Euclidean
Construction	2.94%	2.94%	2.94%	[illegible]
TEC Site	3.85%	3.85%	[illegible]	[illegible]
Parkland 1	0.00%	0.00%	0.00%	[illegible]
High School	0.00%	0.00%	0.00%	[illegible]
Mall	1.61%	1.61%	1.61%	[illegible]
Parkland 2	0.00%	0.00%	0.00%	[illegible]
BareSoil	0.00%	0.00%	0.00%	[illegible]
Fields-A	69.30%	68.10%	[illegible]	[illegible]
Fields-C	68.32%	68.32%	[illegible]	[illegible]
Fields-D	1.89%	0.94%	[illegible]	[illegible]
Grass-A	0.00%	0.00%	[illegible]	[illegible]
Grass-C	3.23%	0.00%	[illegible]	[illegible]
Leaf	15.60%	19.40%	[illegible]	[illegible]
Pine	2.54%	3.82%	[illegible]	[illegible]
Road-A	18.76%	18.56%	[illegible]	[illegible]
Runway-C	0.00%	0.00%	[illegible]	[illegible]
Runway-F	0.00%	0.00%	[illegible]	[illegible]
Swamp-A	15.20%	66.47%	[illegible]	[illegible]
Swamp-B	0.00%	0.00%	[illegible]	[illegible]
Urban-D	0.00%	0.00%	[illegible]	[illegible]
Urban-F	0.00%	0.00%	[illegible]	[illegible]
Urban-I	0.00%	0.00%	[illegible]	[illegible]
Water-A1	0.00%	16.90%	[illegible]	[illegible]
Water-A2	0.00%	2.60%	[illegible]	[illegible]
Water-C	0.00%	15.30%	[illegible]	[illegible]

The Fields-A and Fields-C sites generated the highest [illegible] to be agricultural fields. Their spectral behavior and the [illegible] be understood by referring to Appendix A, which lists the [illegible] May 85, August 85, October 85, and March 85.

Consider the mean spectra for Fields-A. Notice that for the [illegible] as May 85, October 85, and March 85), the mean spectral [illegible] the signature of vegetation, but rather it is closer to that of soil. [illegible] signature changes quite dramatically to one that is indicative of [illegible] This is, of course, quite typical behavior for crops. It also [illegible] where the majority of the Fields-A samples were labeled [illegible] classifier used). In addition, approximately 20 to 26 percent of the [illegible] deciduous trees (For further reference, see Table B4 in [illegible]

Tables 4-8 to 4-12 summarize the results of using the modified Bayes algorithm and invoking three different rejection thresholds. The tightest rejection criterion tested was $\chi^2_{.01}(6) = 16.81$. Samples having a squared Mahalanobis distance (to the class selected by the modified Bayes algorithm) greater than this value are rejected. Of the three thresholds, this value should result in the least number of classification errors, but the most number of rejected/unclassified samples. According to this value where $\chi^2_{.01}(6)$ corresponds to a chi-squared distribution having 6 degrees of freedom and a significance value of $\alpha = .01$, there is theoretically a 1.0 percent chance that the sample in question belonged to the class that was selected, but was rejected. This corresponds to what is commonly called a Type I error.

Decreasing this α number will result in a smaller Type I error; however, it will also lead to a higher Type II error. A Type II error corresponds to accepting the sample as the class that was selected when it actually corresponds to some other class. Increasing the Type II error will, of course, increase the classification error; however, a smaller number of samples will be rejected.

Some conventional software systems (such as LAS) have the capability to invoke a chi-squared threshold, but have a limit whereby the α value cannot be decreased to less than about $\alpha = .005$. At first consideration, it would seem this limitation is most reasonable since theoretically at this value of $\alpha = .005$ we would only be rejecting 0.5 percent of the class population. However, TEC's past experience seemed to indicate that this value may actually be too stringent. Even with such a low value, the result of applying a threshold corresponding to this significance value is that too large a portion of the scene is rejected.

The problem of rejecting too many samples using such a low significance value can be understood if one recalls that scenes have a large amount of natural diversity. The prototypes used for training and the samples that need to be classified may correspond physically to perhaps 80-90 percent of the same material, but small portions of materials are altering the spectral signature.

For this reason, two other threshold values that correspond to five times the $\chi^2_{.01}(6)$ distance and seven times the $\chi^2_{.01}(6)$ distance are also tested. It should be expected that the threshold corresponding to seven times the distance would produce the highest classification error (but the least number unlabeled pixels).

Table 4-8 shows the Bayes auto-classification results for the three threshold distances. Because it has already been established that the modified Bayes had produced no errors for the training samples, the results are simply given as the percentage of unclassified pixels. Notice that $\chi^2_{.01}(6) = 16.81$ rejected a moderate number of samples ([illegible] % of Water; [illegible] % of C Veg; [illegible] % of Concrete; 0.0% of others). The unweighted average rejection percentage over the eight classes is

$$([illegible]) \% = [illegible]\ \%$$

This average rejection of 1.07 percent is very close to the theoretical value of 1.0 percent for $\alpha = 0.01$.

[illegible]

[illegible]

	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]

[illegible]

[illegible]

	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]				
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]		[illegible]	
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	
[illegible]	[illegible]	[illegible]

[illegible]

[illegible]

[illegible]

Table 4-11 Bayes Omission Results Using 5 Times the χ^2 Value - Trial 3

This table lists the percentage of misclassified and unclassified pixels, as well as the total percentage omitted, for each of the test sites in B2 for a threshold value of 84.05, derived from the chi-square statistic with 6 degrees of freedom.

$d < \chi^2_{.01}(6) * 5$

	Misclassified	Unclassified	Omission
Construction	0.00%	41.18%	41.18%
TEC Site	0.00%	100.00%	100.00%
Parkland 1	0.00%	0.00%	0.00%
High School	0.00%	96.43%	96.43%
Mall	0.00%	96.77%	96.77%
Parkland 2	0.00%	14.49%	14.49%
BareSoil	0.00%	100.00%	100.00%
Fields-A	26.40%	65.80%	92.20%
Fields-C	0.00%	100.00%	100.00%
Fields-D	1.89%	0.00%	1.89%
Grass-A	0.00%	60.92%	60.92%
Grass-C	3.23%	3.23%	6.45%
Leaf	15.60%	0.00%	15.60%
Pine	1.53%	1.27%	2.80%
Road-A	1.80%	65.47%	67.27%
Runway-C	0.00%	7.69%	7.69%
Runway-F	0.00%	22.68%	22.68%
Swamp-A	0.00%	57.08%	57.08%
Swamp-B	0.00%	50.00%	50.00%
Urban-D	0.00%	62.96%	62.96%
Urban-F	0.00%	93.33%	93.33%
Urban-I	0.00%	7.14%	7.14%
Water-A1	0.00%	0.00%	0.00%
Water-A2	0.00%	0.00%	0.00%
Water-C	0.00%	100.00%	100.00%

Percentage of test set unclassified = 27.59%

Table 4-12 Bayes Omission Results Using 7 Times the χ^2 Value - Trial 3

This table lists the percentage of misclassified and unclassified pixels, as well as the total percentage omitted, for each of the test sites in B2 for a threshold value of 117.67, derived from the chi-square statistic with 6 degrees of freedom.

$d^2 < \chi^2_{.01}(6) * 7$

	Misclassified	Unclassified	Omission
Construction	0.00%	17.65%	17.65%
TEC Site	0.00%	100.00%	100.00%
Parkland 1	0.00%	0.00%	0.00%
High School	0.00%	85.71%	85.71%
Mall	0.00%	87.10%	87.10%
Parkland 2	0.00%	1.45%	1.45%
BareSoil	0.00%	100.00%	100.00%
Fields-A	31.10%	57.70%	88.80%
Fields-C	7.92%	92.08%	100.00%
Fields-D	1.89%	0.00%	1.89%
Grass-A	0.00%	39.08%	39.08%
Grass-C	3.23%	3.23%	6.45%
Leaf	15.60%	0.00%	15.60%
Pine	2.29%	0.25%	2.54%
Road-A	2.59%	54.09%	56.69%
Runway-C	0.00%	1.28%	1.28%
Runway-F	0.00%	13.40%	13.40%
Swamp-A	2.38%	36.96%	39.34%
Swamp-B	0.00%	8.33%	8.33%
Urban-D	0.00%	3.70%	3.70%
Urban-F	0.00%	86.67%	86.67%
Urban-I	0.00%	0.00%	0.00%
Water-A1	0.00%	0.00%	0.00%
Water-A2	0.00%	0.00%	0.00%
Water-C	0.00%	100.00%	100.00%

Percentage of test set unclassified = 21.91%

4.5 Results of Trial 4

Trial 4 investigates the effect of reducing the number of bands, repeating the analysis that was done on the modified Bayes approach in Trial 3 using the four Landsat TM bands B3, B4, B5, B7, rather than all six bands. Notice that the chi-square distance threshold value changes because degrees of freedom for the distribution change from six to four. However, for consistency they were selected in the same manner: one times the chi-square distance, five times the chi-square distance, and seven times the chi-square distance.

Table 4-13 shows the auto-classification results using only B3, B4, B5, and B7. The auto-classification of 4 bands produced almost the same low error rate at $\chi^2_{.01}(4)$ as that of 6 bands, except that the Grass-B class contained 4.17 percent error (compared to 0.0% for 6 bands). The results for the other chi-squared values were 0.00 percent for all classes (identical to the results achieved for 6 bands.

Table 4-14 shows the commission results for these four bands. The same trend of decreasing errors for decreasing thresholds is seen. Except for the lowest threshold value $\chi^2_{.01}(4) = 13.28$, the results are almost the same. For the lowest threshold, however, 81 errors occur for the Grass-B class using 4 bands vs. 39 errors using 6 bands. Referencing the contingency table, the classifier is calling 80 of these 81 errors Grass-B, when they should have been called D. Veg.

Based on these results, there would seem to be little impact to reducing the bands. However, the omission error results, listed in Tables 4-15 to 4-17, show some problems. As was the case for 6 bands, the trial for the lowest chi-squared threshold, while maintaining a low misclassification error, resulted in mostly unclassified data. Proceeding to the next highest threshold of $\chi^2_{.01}(4) \cdot 5 = 66.4$, more of the data was classified. Unfortunately, a large number were misclassified. Referring to the contingency results in Appendix B, some of the degradation in going from 6 bands to 4 bands (for this threshold) can be compared as follows:

CLASS	6-band error	4-band error	Major cause of Problem
TEC Site	0.00%	19.23%	Samples being labeled as B. Roof
High School	0.00%	60.71%	Samples being labeled as B. Roof
Mall	0.00%	62.90%	Samples being labeled as B. Roof
BareSoil	0.00%	73.68%	Samples being labeled as B. Roof
Fields-A	26.40%	54.50%	Samples being labeled as B. Roof
Road-A	1.80%	33.53%	Samples being labeled as B. Roof

Apparently, the reduction in the number of bands causes confusion between the samples containing soil and/or concrete and are being confused with the Bright Roof class, that is believed (but not yet confirmed) to be metal. There does not seem to be a problem in confusing vegetation; however, mixtures of soil and vegetation such as Fields-A were also confused with this Bright Roof class.

Based on these results, the reduction of bands from 6 to 4 cannot be recommended. Further reduction beyond 4 bands is highly discouraged.

Table 4-13 Bayes Auto-Classification Results Using χ^2 Thresholds - Image [illegible] (only bands B3, B4, B6, B7 were used)

[illegible]

	χ^2 [illegible]	χ^2 [illegible]	χ^2 [illegible]
Water 1	[illegible]	[illegible]	[illegible]
S. Sand	[illegible]	[illegible]	[illegible]
B. Veg	[illegible]	[illegible]	[illegible]
C. Veg	1.67%	[illegible]	[illegible]
Asphalt	[illegible]	[illegible]	[illegible]
Concrete	[illegible]	[illegible]	[illegible]
Water 2	[illegible]	[illegible]	[illegible]
Grass-B	[illegible]	[illegible]	[illegible]

Table 4-14 Bayes Commission Results Using χ^2 Thresholds - Image [illegible] (only bands B3, B4, B6, B7 were used)

[illegible]

	χ^2 [illegible]	χ^2 [illegible]	χ^2 [illegible]
Water 1	[illegible]	[illegible]	[illegible]
S. Sand			
B. Veg	[illegible]	[illegible]	[illegible]
C. Veg	[illegible]	[illegible]	[illegible]
Asphalt	[illegible]	[illegible]	[illegible]
Concrete	[illegible]	[illegible]	[illegible]
Water 2	[illegible]	[illegible]	[illegible]
Grass-B	[illegible]	[illegible]	[illegible]

	χ^2 [illegible]	χ^2 [illegible]	χ^2 [illegible]
Water 1	[illegible]	[illegible]	[illegible]
S. Sand			
B. Veg	[illegible]	[illegible]	[illegible]
C. Veg	[illegible]	[illegible]	[illegible]
Asphalt	[illegible]	[illegible]	[illegible]
Concrete	[illegible]	[illegible]	[illegible]
Water 2	[illegible]	[illegible]	[illegible]
Grass-B	[illegible]	[illegible]	[illegible]
TOTAL	[illegible]	[illegible]	[illegible]

4.6 Results of Trial 5

The objective of this trial was to investigate the behavior of the three well-known supervised classifiers -- the Standard Bayes discriminant, the Mahalanobis distance, and the minimum Euclidean distance -- on data acquired over different seasons and years. Because of the desire to proceed with testing the linear mixture modeling, the modified Bayes discriminant using a minimum variance criterion and/or chi-squared threshold was not tested. The classifiers' performance was tested against their own training data (auto-classification), and the ground truth (GT) test data extracted from the imagery. Data from the five mosaic datasets were used: May 1987, May 1985, August 1985, October 1985, and March 1989. Therefore, the effect of different seasons for the same year could be studied, as well as the effect of the same season for different years.

Results and discussion of the auto-classification analysis are first presented, followed by results and discussion of classification analysis on the ground truth data (GT). A description of the mosaic data sets, and the training set acquisition process and properties were discussed previously in Sections 3.1 and 3.2, respectively. Training statistics (mean vectors and covariance matrices) are listed in Appendix A. More detailed results for the auto-classification runs are given in Appendix C.

Auto-Classification Analysis of Training Areas - Set B2

Auto-classification runs were made on Training Set B2 to test the performance of the Bayesian, Mahalanobis, and Euclidean classifiers when applied to its own training data. These runs were repeated using data from *all five mosaic images:* May 1987, May 1985, August 1985, October 1985, and March 1989. Training Set B2 consists of the 20 classes numbered 100-194, as shown in Table 3-3 (Section 3.2). During this trial, classes 8-13 were not used.

The performance of these classifiers is summarized in Table 4-18. This table shows the percentage of correct hits for each class for all three methods and also the average of correct hits for each method (where each class is weighted equally). Note that this summary consolidates the results of the 20 training classes into 16 classes by combining the three *field* classes (Fields-A, Fields-B, And Fields-C) into a class called Fields, and combining the three *water* classes (Water-A1, Water-A2, And Water-C) into a class called Water. Appendix C contains a table showing the results without the consolidation.

The results are reported with this consolidation because we did not want to penalize the classifiers for confusion between similar classes that would eventually be consolidated by subsequent operations. We could have similarly combined many of the others (such as road and runway); however, the performance was so good it did not seem necessary, and in addition, the ability of the classifiers to maintain separability between such fine classes provides additional insight into their behavior.

The Bayesian discriminant classifier proved to be the best of the three methods. The Bayesian results were consistent across all five dates tested. The overall performance, as well as the performances of all individual cases, was excellent. By consolidating only field classes and water classes, the average percentages of error were **1.95%, 1.27%, 0.72%, 2.82% and 3.68%** for May 87, May 85, August 85, Oct 85, and March 89, respectively. The highest error for any individual class occurred in the March 89 data for Leaf and had a value of 11.20 percent.

The second best classifier proved to be the Mahalanobis distance classifier. Generally, the performance was very good, with most errors below 10.0 percent. The average percentages of error were **5.35%, 2.81%, 0.96%, 4.83%, and 11.03%** for the five dates. However, the consistency between dates was not as good. For example, the Grass-B class maintained an error rate of less than 10.0 percent for all dates except March 89, for which it increased to **50%**. The corresponding contingency table (not shown) reported that 41.67 % (10 out of 24 samples) of the Grass-B samples were incorrectly labeled as Leaf. Two other relatively poor performers for this March 1989 data were Grass-A at **21.84%** and Grass-C at **32.26%**; however, they are not as bad as they seem. The Mahalanobis classifier (incorrectly) labeled 20.69% (18 out of 87) of the Grass-A samples, and 32.26% (10 out of 31) as Fields-A. If the Grass-A and Fields-A were later consolidated, the 87 Grass-A samples would have a **1.5%** error rate. If the Grass-C and Fields-A were later consolidated, the 31 Grass-C samples would have a **0.0%** error rate.

Although not as good as the above two methods, the Euclidean distance classifier provided very good results, although somewhat lower and less consistent. The average percentages of error were **13.20%, 8.19%, 4.64%, 14.56%, and 21.59%** for the five dates. Again consistency among dates and individual cases was not as good as for the Bayesian method.

Table 4-18 Auto-Classification Summary for Training Set B.

Field Classes Combined and Water Classes Combined

	Training Data MY87_1000Samples			Training Data MY85_1000Samples		
	Bayes	Mahalanobis	Euclidean	Bayes	Mahalanobis	Euclidean
Baresoil	0.00%	0.00%	0.00%	0.00%	0.00%	2.63%
Fields	8.70%	2.49%	63.88%	7.95%	2.65%	39.27%
Grass-A	2.30%	5.75%	2.30%	1.15%	12.64%	1.15%
Grass-B	0.00%	8.33%	16.67%	0.00%	0.00%	12.50%
Grass-C	0.00%	16.13%	16.13%	0.00%	3.23%	6.45%
Leaf	3.10%	27.30%	8.20%	1.30%	1.80%	6.50%
Pine	2.80%	8.14%	10.69%	2.80%	12.72%	17.56%
Road-A	8.38%	10.98%	30.34%	3.99%	6.99%	19.36%
Runway-C	5.13%	5.13%	5.13%	1.28%	1.28%	0.00%
Runway-F	0.00%	0.00%	4.12%	0.00%	0.00%	0.00%
Swamp-A	0.45%	0.30%	30.40%	1.34%	3.13%	9.84%
Swamp-B	0.00%	0.00%	16.67%	0.00%	0.00%	8.33%
Urban-D	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Urban-F	0.00%	0.00%	6.67%	0.00%	0.00%	0.00%
Urban-I	0.00%	0.00%	0.00%	0.00%	0.00%	7.14%
Water	0.40%	0.99%	0.05%	0.50%	0.55%	0.35%
Average	**1.95%**	**5.35%**	**13.20%**	**1.27%**	**2.81%**	**8.19%**

Table 4-18 Auto-Classification Summary for Training Set B (continued).

Field Classes Combined and Water Classes Combined

	Training Data AG85_1000Samples			Training Data [illegible]		
	Bayes	Mahalanobis	Euclidean	Bayes	Mahalanobis	Euclidean
Baresoil	0.00%	0.00%	2.63%	[illegible]	[illegible]	[illegible]
Fields	2.98%	2.73%	18.23%	[illegible]	[illegible]	[illegible]
Grass-A	0.00%	0.00%	0.00%	[illegible]	[illegible]	[illegible]
Grass-B	0.00%	0.00%	0.00%	[illegible]	[illegible]	[illegible]
Grass-C	0.00%	0.00%	0.00%	[illegible]	[illegible]	[illegible]
Leaf	3.00%	2.60%	9.10%	[illegible]	[illegible]	[illegible]
Pine	2.80%	6.62%	11.70%	[illegible]	[illegible]	[illegible]
Road-A	1.80%	0.80%	[illegible]	[illegible]	[illegible]	[illegible]
Runway-C	0.00%	1.28%	[illegible]	[illegible]	[illegible]	[illegible]
Runway-F	0.00%	0.00%	0.00%	[illegible]	[illegible]	[illegible]
Swamp-A	0.45%	0.75%	[illegible]	[illegible]	[illegible]	[illegible]
Swamp-B	0.00%	0.00%	0.00%	[illegible]	[illegible]	[illegible]
Urban-D	0.00%	0.00%	0.00%	[illegible]	[illegible]	[illegible]
Urban-F	0.00%	0.00%	6.67%	[illegible]	[illegible]	[illegible]
Urban-I	0.00%	0.00%	[illegible]	[illegible]	[illegible]	[illegible]
Water	0.55%	0.60%	[illegible]	[illegible]	[illegible]	[illegible]
Average	**0.72%**	**0.96%**	**4.64%**	[illegible]	[illegible]	[illegible]

Field Classes Combined and Water Classes Combined

	Training Data [illegible]		
	Bayes	Mahalanobis	Euclidean
Baresoil	0.00%	[illegible]	[illegible]
Fields	8.12%	[illegible]	[illegible]
Grass-A	2.30%	[illegible]	[illegible]
Grass-B	0.00%	[illegible]	[illegible]
Grass-C	6.45%	[illegible]	[illegible]
Leaf	11.20%	[illegible]	[illegible]
Pine	[illegible]	[illegible]	[illegible]
Road-A	[illegible]	[illegible]	[illegible]
Runway-C	[illegible]	[illegible]	[illegible]
Runway-F	[illegible]	[illegible]	[illegible]
Swamp-A	[illegible]	[illegible]	[illegible]
Swamp-B	0.00%	[illegible]	[illegible]
Urban-D	0.00%	[illegible]	[illegible]
Urban-F	0.00%	[illegible]	[illegible]
Urban-I	0.00%	[illegible]	[illegible]
Water	[illegible]	[illegible]	[illegible]
Average	[illegible]	[illegible]	[illegible]

Classification Analysis of Ground Truth Areas - Set GT

Using the 20 training classes just discussed (Classes 100-194), classification runs were made for all 5 dates on the ground truth test data (Set GT). Because the modified Bayesian technique has only been implemented experimentally on a microcomputer workstation (discussed in Section 2.3), and Trial 5 was conducted separately on the LAS software, only the standard Bayes and Euclidean minimum distance methods were tested. Although the modified Bayesian could have easily been tested on the five dates, testing of the linear mixture model was a higher priority.

Tables 4-19 and 4-20 list the commission and omission results, respectively. In general, the results are not consistent across dates and there are wide swings in performance for most classes, particularly, for vegetation-related classes (Grasses/Fields, Swamp, Leaf and Pine).

Although the Bayesian results were usually better than the Euclidean distance, they were not consistently better across dates. For example, consider the omission results for Grass/Fields. The Bayesian classifier performed better in May 1985, May 1987, and March 1989, however, the Euclidean distance performed better in AG85 and OC85. Now consider the commission results for Grass. The Bayesian classifier performed better in May 1987 and March 1989, but worse on the other dates.

The Bayesian classifer performed consistently better in omission errors for the Urban class, however, the consistency did not hold for commission errors. In fact, the only exception to inconsistency is the water class where the Euclidean distance performed better on all dates.

Given the theoretical advantages of using the Bayesian method as discussed in Section 2.1, and the success of Trial 3 in improving the standard Bayesian classifier results by invoking a minimum variance criterion and the chi-squared rejection criteria, this method should be preferred over the Euclidean minimum distance method. Although the latter performed consistently better at water during Trial 5, recall that Trial 3 demonstrated a dramatic improvement in the modified Bayesian method for detecting water.

The best of the 5 dates for detecting Swamp was OC85 (although the commission error remained high). This result should be taken with caution, however, because only one swamp site for ground truth was used. There are, of course, a wide variation of swamp spectra, corresponding to various proportions of water and vegetation, as well as various types and vigour of vegetation.

Because no trend is apparent, no definite conclusion can be reached regarding the best time of year, except perhaps to say that August seemed to be the worst performer. However, the difficulty with August might have just been a problem with haze, which was noticeably nonhomogeneous across the scene for this date, and not with the underlying scene content or spectra.

Table 4-19 [illegible]

[illegible]

	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

[illegible]

	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

[illegible]

	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

[illegible]

[illegible]

	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

[illegible]

	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

[illegible]

	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

4.7 Results of Linear Mixing Model Trial

The mixture trials test the utility of using the linear approach discussed in Sections 2.1.5 and 3.5 to uniquely model swamps, which is presumably a mixture of vegetation and water. Ten samples were selected to test the linear mixing model. These were extracted from Dataset B2, and Datset C, and are identified as follows:

Label	Material
C174	Swamp
C175	Swamp
C176	Swamp
C123	Grass
C125	Grass
C133	Leaf
B140	Pine
B160	Asphalt RW
B162	Concrete
B190	Water

The swamps, C174, C175, C176, are the material mixtures being tested. The remaining materials are tested as possible endmembers for the mixtures.

The domain limits defined by each of the pairs of endmembers were determined using the sample mean vectors for each of the endmembers. As mentioned in Section 3.5, these limits must necessarily be considered approximate because each sample is a cloud of data and there are obviously individual endmembers in each sample on the outer portions of the cloud that would increase the width of the domain/interval. The most suitable endmembers, according to the first physical constraint, are those for which one endmember response is lower than the mixture and the other endmember response is higher than the mixture (i.e. the endmember spectra surround the mixture, one above and the other below).

Recall from Section 2.1.5, Figure 1 displays a graph of an idealized signature generated by a 40/60 linear mix of asphalt and concrete. The mixture spectra was in the middle, with the asphalt spectra on the bottom and the concrete spectra on the top. This situation was, of course, fully compliant with the first physical constraint by construction.

Figure 13 displays some of the spectra under study during this trial. One endmember combination (Asphalt B160, Concrete B162) is clearly not compliant with the first constraint. Both spectra lie completely above the Swamp C174 spectra. Another endmember combination (Deciduous C133, Water B190) is mostly compliant. In this case, the swamp is bounded by the endmember spectra, except in band B3 where both endmember responses are below the swamp response. However, the violation is very slight and can probably be accepted if the variance of the features for B3 is considered.

Notice another phenomena occurring for the (Deciduous C133, Water B190) endmember combination. For B1 and B2, the Deciduous response is below the swamp and the Water response is above the swamp. For B4 to B7, the Deciduous response is above and the Water response is below. That is, there is a flip that pivots about some point between B2 and B4. This crossover of the spectra should not be troublesome to the reader, since the physical constraints are still satisfied (the mixture spectra is still bounded by the two endmember spectra).

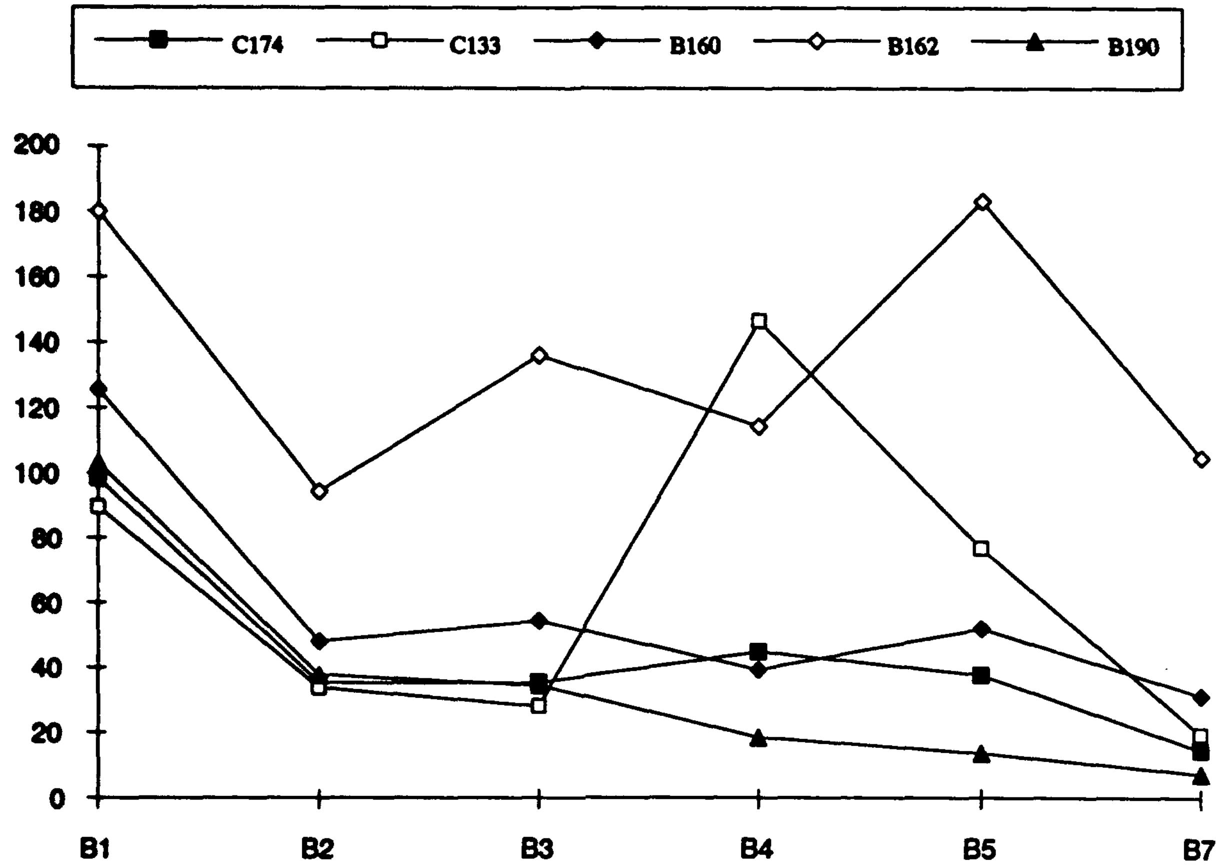


Figure 13. Observed Spectra of Swamp and Candidate Endmembers

Table 4-21 lists the domain limits for some of the endmember combinations. In this table, the mixture (Swamp C174) is placed in the middle of two endmembers. For the endmembers to be completely compliant with the first physical constraint, the Swamp response must lie within the interval defined by the endmember pair for all bands.

Table 4-22 lists the full regression results for one of the endmember models of Swamp C174. Note that both a model with a constant term and without a constant term was generated. This approach is used for all the various combinations. For each combination, the model with a constant is generated. If the constant is found insignificant, it is dropped. For the model to be physically appropriate this must indeed be true. As it turns out, the constant was found to be insignificant in almost all the cases. The detailed regression results are listed in Appendix D. Although only a few examples of the models with a constant are listed, they were indeed tested, and the constants were found to be insignificant.

Regression models are computed with diagnostic statistics for each of the pairwise endmember combinations. An F-ratio is used to assess the statistical significance of a model. If none of the candidate endmember pairs had produced a statistically significant model, then the model would have been expanded to include additional endmembers (up to a 4-component model). However, all the trials produced statistically significant pairwise models.

Table 4-21 Pairwise Domain Limits Surrounding Swamp

MY85	Water B190	Swamp C174	Deciduous C133	Comments on Domain Limits
B1	103.27	98.4	89.95	
B2	37.91	35.439	33.884	
B3	34.86	35.709	28.134	Slightly Outside Interval
B4	19	44.81	146.475	
B5	13.72	37.624	77.442	
B7	7.47	14.984	19.439	

MY85	Water B190	Swamp C174	Concrete B162	
B1	103.27	98.4	180.42	Outside Interval
B2	37.91	35.439	94.74	Slightly Outside Interval
B3	34.86	35.709	136.23	
B4	19	44.81	114.43	
B5	13.72	37.624	182.9	
B7	7.47	14.984	104.94	

MY85	Water B190	Swamp C174	Grass C125	
B1	103.27	98.4	103.794	Outside Interval
B2	37.91	35.439	42.265	Slightly Outside Interval
B3	34.86	35.709	38.735	
B4	19	44.81	149.794	
B5	13.72	37.624	114.853	
B7	7.47	14.984	36.618	

MY85	Water B190	Swamp C174	Asphalt B160	
B1	103.27	98.4	126.04	Outside Interval
B2	37.91	35.439	48.49	Slightly Outside Interval
B3	34.86	35.709	54.86	
B4	19	44.81	39.62	Outside Interval
B5	13.72	37.624	52.12	
B7	7.47	14.984	31.56	

MY85	Asphalt B160	Swamp C174	Deciduous C133	
B1	126.04	98.4	89.95	
B2	48.49	35.439	33.884	
B3	54.86	35.709	28.134	
B4	39.62	44.81	146.475	
B5	52.12	37.624	77.442	Significantly Outside Interval
B7	31.56	14.984	19.439	Outside Interval

Table 4-22 Regression Results for One of the Endmember Models of Swamp

This table shows the regression results and analysis of variance (ANOVA) tables for a Linear Model of Swamp C174 that is comprised of a mixture of Leaf C133 and Water B190. The results are generated for a linear model both with and without a constant.

DEP VAR: Swamp C174 N: 6 MULTIPLE R: 0.990 SQUARED MULTIPLE R: 0.981
ADJUSTED SQUARED MULTIPLE R: 0.968 STANDARD ERROR OF ESTIMATE: 5.027

VARIABLE	COEFFICIENT	STD ERROR	STD COEF	TOLERANCE	T	P(2 TAIL)
CONSTANT	5.993	4.123	0.000	.	1.454	0.242
Leaf C133	0.196	0.047	0.337	0.977	4.184	0.025
Water B190	0.710	0.065	0.881	0.977	10.934	0.002

ANALYSIS OF VARIANCE

SOURCE	SUM-OF-SQUARES	DF	MEAN-SQUARE	F-RATIO	P
REGRESSION	3907.349	2	1953.674	77.308	0.003
RESIDUAL	75.814	3	25.271		

MODEL CONTAINS NO CONSTANT.

DEP VAR: Swamp C174 N: 6 MULTIPLE R: 0.996 SQUARED MULTIPLE R: 0.992
ADJUSTED SQUARED MULTIPLE R: 0.990 STANDARD ERROR OF ESTIMATE: 5.684

VARIABLE	COEFFICIENT	STD ERROR	STD COEF	TOLERANCE	T	P(2 TAIL)
Leaf C133	0.241	0.040	0.372	0.542	6.066	0.004
Water B190	0.753	0.065	0.706	0.542	11.508	0.000

ANALYSIS OF VARIANCE

SOURCE	SUM-OF-SQUARES	DF	MEAN-SQUARE	F-RATIO	P
REGRESSION	15732.427	2	7866.214	243.517	0.000
RESIDUAL	129.210	4	32.302		

In tables 4-23 to 4-26, the [illegible] in Appendix D [illegible], [illegible] criteria used to evaluate the models.

The selection [illegible] four criteria.

1. [illegible] physical [illegible].

2. Any [illegible] F-ratio [illegible]

3. The [illegible] are positive and [illegible]

4. Each and every [illegible]

A degree of compliance (DOC) [illegible]

DOC = 3 [illegible]

DOC = 2 [illegible]

DOC = 1 Partially [illegible]

DOC = 0 No compliance. [illegible]

As just mentioned in the second selection criteria, for the model [illegible] statistically significant, the F ratio must be greater than a certain threshold. This threshold [illegible] the F statistic.

$$F_{thresh} = F_{k,\,n-k-1}(\alpha)$$

In our case, k = 2 (the number of [illegible]), [illegible] (the number of [illegible]), and we [illegible] the significance level to be α = .01. Therefore, the threshold becomes

$$F_{thresh} = F_{[illegible]}(.01) = 30.82$$

Observe the F ratio results in Tables 4-23 to 4-[illegible], and notice that except in one case, the models using a constant and the [illegible] pair [illegible], all the models are statistically significant. Once the constant for [illegible] pair is dropped, there are no exceptions. All the models shown are statistically significant. [illegible] models are definitely more significant than others. However, unless [illegible] imposed, it is important to realize that all the models are statistically [illegible], even though we will still [illegible] the models with the largest F ratios. Another immediate [illegible] is that [illegible] necessary to expand the [illegible] model to include additional [illegible].

[illegible]

[illegible]

[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]		[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]		[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]		[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]		[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]		[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

[illegible]

[illegible]

[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]		[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]		[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]			[illegible]	[illegible]		[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

[illegible]

[illegible]

[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]		[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]		[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]		[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]		[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]		[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]		[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

$$[illegible]$$

Table 4-26 Diagnostics for Candidate Mixtures to Model Swamp C174

RESIDUALS

	B1	B2	B3	B4	B5	B7
Leaf C128, Water B190	-1.86	-1.28	2.68	-4.77	8.64	4.68
Concrete B162, Water B190	2.91	-4.28	-4.29	14.71	0.63	-5.97
GreenC128, Water B190	0.40	-1.60	1.68	-2.09	2.29	1.45
Grass C125, Water B190	1.82	-2.53	-1.50	5.15	-2.78	-2.30
Asphalt B160, Water B190	0.50	-2.86	-5.92	15.34	-0.03	-7.73
Pine B140, Water B190	-0.32	-1.14	2.34	-3.28	4.16	2.68
Pine B140, Asphalt B160	5.26	-4.17	-2.76	1.59	-4.33	-7.47

LEVERAGE

	B1	B2	B3	B4	B5	B7
Leaf C128, Water B190	0.80	0.11	0.08	0.79	0.20	0.01
Concrete B162, Water B190	0.85	0.10	0.16	0.14	0.56	0.19
GreenC128, Water B190	0.81	0.11	0.08	0.61	0.36	0.03
Grass C125, Water B190	0.82	0.10	0.09	0.35	0.56	0.08
Asphalt B160, Water B190	0.82	0.10	0.12	0.13	0.58	0.24
Pine B140, Water B190	0.80	0.11	0.09	0.67	0.31	0.02
Pine B140, Asphalt B160	0.65	0.10	0.17	0.85	0.18	0.05

COOK'S DISTANCE

	B1	B2	B3	B4	B5	B7
Leaf C128, Water B190	0.33	0.00	0.01	6.39	0.37	0.00
Concrete B162, Water B190	6.92	0.01	0.09	0.25	0.01	0.06
GreenC128, Water B190	0.00	0.04	0.04	2.00	0.55	0.01
Grass C125, Water B190	1.08	0.03	0.01	0.89	0.92	0.02
Asphalt B160, Water B190	0.04	0.00	0.03	0.24	0.00	0.14
Pine B140, Water B190	0.25	0.01	0.03	3.13	0.53	0.01
Pine B140, Asphalt B160	2.10	0.00	0.03	1.40	0.16	0.05

where e_i is the ith residual, T is a threshold that defines "small", and N = 6 (the number of bands). Of course, the definition of small is a bit arbitrary.

Based on the residual results in Table 4-26, if we define T = 5, then we are left with two solutions: {GreenC128, Water B190} and {Pine B140, Water B190}. If we define T = 3, then we are left with a single solution: {GreenC128, Water B190}.

Leverage and Cook's Distance are measures of influence. If leverage for a point is greater than $2p/N = .66$, or if Cook's distance is greater than about d = 1, then that point is considered influential. Table 4-26 indicates that bands B1 and B4 are influential for all the endmember combinations. Although these measures are interesting in that they convey this property, it does not appear valuable in identifying good or poor endmember combinations. The measures do indicate, however, that bands B1 and B4 are more influential than the other bands in determining the model.

5.0 CONCLUSIONS

5.1 Conclusions Regarding the Graphical Analysis

The graphical analysis indicated a possible degeneracy in the spectral space defined by broad-band spectral sensors (such as Landsat TM), where a mixture of materials could combine to form a signature identical to the signature of certain pure pixels. In particular, coniferous and deciduous trees were observed to lie in a region of spectral space occupied by certain mixtures of water and vegetation (e.g. certain types of swamp). For such situations, no algorithm, regardless of its complexity, will separate such classes. The spectral information just simply doesn't exist to distinguish them. This provides motivation for using narrow-band spectral imagery, consisting of higher spectral resolution and more bands.

The addition of more spectral bands with increased spectral resolution, hopefully, can eliminate the degenerate cases. However, there is no guarantee that this approach will be successful. The underlying spectra might be quite bland and not contain distinguishing absorption features. Therefore, incorporating such data, although more volumous, would not necessarily provide increased spectral information.

5.2 Conclusions Regarding the Spectral Classification

Performance of the conventional classifiers as typically applied to Landsat TM is unacceptable for the general application of extracting natural and manmade features. The most disturbing behavior of the conventional Bayesian and Mahalanobis classifiers was the tendency to mislabel water and marsh/swamp features in a scene as asphalt. This type of error has serious consequences to military and environmental applications (e.g. A convoy of jeeps and trucks would prefer to stay on the roads and not drive into a swamp). In this regard, the Euclidean classifier performed much better.

The Euclidean minimum distance classifier performed better at not mislabeling water features. However, it did not perform as well as the Bayesian or Mahalanobis classifier for many other features.

In many cases, the problems found with the conventional classifiers were not due to a lack of spectral separability between materials or a lack of spectral resolution. The problem was often one (or a combination) of the following:

a. Correspondence between the objects of interest in a scene and the materials (the classifier is classifying the materials, not the objects).

b. Correspondence between the samples in the scene and the available prototype classes because there is an insufficient number of prototype classes.

c. Samples in the scene are mixtures of materials represented by the prototype classes.

d. Difficulties with covariance matrices modeling the spectral variance of certain classes, particularly, water.

The performance of the Bayesian and Mahalanobis classifier was improved to an acceptable level by using a minimum variance criterion on class covariances and a chi-squared rejection criterion.

a. The use of a minimum variance criterion was shown to correct the problems associated with modeling the spectral variance of water.

b. The use of a chi-squared rejection allowed samples that did not correspond to a prototype class or that correspond to a mixture of classes to be rejected. The error rate was reduced dramatically, labeling as unknown those samples that were previously misclassified.

c. The chi-squared rejection criterion, as sometimes implemented on other systems, is not acceptable. Often times, there is the need to allow a larger rejection distance than what is available. The software written during this effort allows the use of such larger rejection distances.

The chi-squared rejection criterion would be particularly useful for targeting applications. An analyst could train on a specific ground feature of interest. By invoking a tight threshold distance, the analyst would have a very high degree of confidence that any ground feature identified as the target material was indeed classified correctly.

Reducing the number of Landsat TM bands from six to four, significantly increased both commission and omission classification errors.

Clearly, more work needs to be done in studying the effect of season and year on classification performance. The existing multidate/multiscene montage data are in a suitable form to study this effect since numerous training, test, and ground truth sites have been extracted. However, the task was beyond the level of effort that could be allocated. Other technical issues have presented themselves that should be addressed first.

In particular, the lack of consistency and wide swings in performance for the Euclidean minimum distance and conventional Bayesian classifier suggest some fundamental instabilities. Two candidate sources are (1) inadequate estimates of the class covariance matrices introduced by quantization effects and outliers in the training samples, and (2) violations of model assumptions and possible degeneracies in the spectral space introduced by mixtures as well as changes in mixing proportions of aggregate materials (e.g. swamps).

The modified Bayes approach has taken some steps to overcome these problems. The minimum variance criterion seems to have corrected the problem of quantization effects (small variance) on the covariance estimates, and the chi-squared rejection threshold flags potential mixture candidates. Therefore, what remains is to incorporate a mechanism for reducing the effect of outliers on the covariance estimates, and a method to handle mixtures.

The experience gained in this effort should be useful to future spectral sensing work involving higher spectral resolution data. In particular, the variance of spectral components is likely to have an adverse effect on any algorithm that does not appropriately incorporate this phenomena. For example, it becomes quite clear from observing the signatures of various grass sites that there is no unique grass signature. Similarly, there is no unique water signature; no unique field signature; no unique asphalt signature; etc. Unless one is looking for unique absorption features of a specific material, it will become necessary to incorporate variance. If a reference library of spectral data is used in the processing, the spectral variance of materials must be incorporated in or be computable from the library.

Also remember that many of the classification errors occurred because either the samples in question did not correspond to a prototype class, or they were mixtures of the materials represented

by the prototype classes. The lack of spectral [illegible] the prototype classes were shown to be [illegible] classifying this data. If one makes the [illegible] multispectral imagery and a reference library of [illegible] likely to be the same. Misclassifications are likely [illegible] are not adequately represented in the spectral library [illegible] spectral library.

5.3 Conclusions Regarding the [illegible]

The linear mixing model was successful only [illegible] The trials showed the basic model [illegible] were mathematically legitimate, but physically [illegible]

Invoking the two proposed physical [illegible] case of swamps, only models containing [illegible] residual threshold, it was possible to [illegible] studies will be necessary to determine the [illegible] candidates, in general, will reduce to a single [illegible] the number of multiple solutions, not eliminated them.

Regarding the first physical constraint, [illegible] been to choose the extremums of the [illegible] extremums were not outliers. This [illegible] increased the complexity of implementing the [illegible] However, considering the reasonable amount of [illegible] in the future.

Increasing the number and resolution of the [illegible] any existing absorption bands to eliminate [illegible] large residual errors near the absorption [illegible] continuous spectra of the endmembers [illegible] resolved bands might not provide any more [illegible] correlation of bands, inherent in [illegible]

Statistical tests conducted on the [illegible] pairwise model to include additional [illegible] endmembers is discouraged.

Only one category of mixture was tested [illegible] for other categories of mixtures.

6.0 REFERENCES

Anderson, T. W. *An Introduction to Multivariate Statistical Analysis*. 2nd Edition, New York, NY: John Wiley & Sons, 1984, p235.

Bow, Sing-Tze. *Pattern Recognition - Applications to Large Data-Set Problems*. New York, NY: Marcel Dekker, Inc., 1984.

Gillespie, A.R., Smith, M.O., Adams, J.B., Willis, S.C., Fischer, A.F., and Sabol D.E. *Interpretation of Residual Images: Spectral Mixture Analysis of AVIRIS Images*. Proceedings of the Second Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) Workshop, JPL Publication 90-54, June 1990.

Johnson, Richard A., Wichern, Dean W. *Applied Multivariate Statistics*. 2nd Edition, Englewood Cliffs, NJ: Prentice-Hall, 1988, p 493 and p513.

Montgomery, Douglas C. and Peck, Elizabeth A. *Introduction to Linear Regression Analysis*. 2nd Edition, New York, NY: John Wiley & Sons, 1992, Chapter 4.

Roberts, D.A., Smith, M.O., and Adams, J.B. Leaf Spectral Types, Residuals, and Canopy Shade in an AVIRIS Image. Proceedings of the Third Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) Workshop, JPL Publication 91-28, May 1991.

Rand, Robert S. *A Hybrid Methodology for Detecting Cartographically Significant Features Using Landsat TM Imagery*. Fort Belvoir, VA: U.S. Army Topographic Engineer Center, ETL-0589, September 1991.

Rand, Robert S., Davis, Donald A., Satterwhite M.B., and Anderson, John E. *Methods of Monitoring the Persian Gulf Oil Spill Using Digital and Hardcopy Multiband Data*. Fort Belvoir, VA: U.S. Army Topographic Engineer Center, TEC-0014, August 1992.

Rand, Robert S. and Davis, Donald A. *Semi-Automated Demonstration of Techniques to Expedite Production of Tactical Terrain Analysis Data Bases Using Landsat TM*. Fort Belvoir, VA: U.S. Army Topographic Engineer Center, internal report.

Searle, S.R. *Linear Models*. New York, NY: John Wiley & Sons, 1971, Chapter 3.

Sheffield, Charles and Richardson, Gil . *New Methods of Change Detection Using Multispectral Data*. Fort Belvoir, VA: U.S. Army Topographic Engineer Center, May 1991. Work performed by Earth Satellite Corporation, Rockville MD 20852.

Therrien, Charles W. *Decision Estimation and Classification*. New York, NY: John Wiley & Sons, 1989.

Wickham, James D. *Land Cover Change Mapping Using Landsat Thematic Mapper Data*. Chevy Chase, MD: Earth Satellite Corporation, July 1988. Work performed under TEC contract DACA76-86-C-0018.

APPENDIX A: Supporting Statistical Data

This appendix contains mean vectors, covariance matrices, and correlation matrices for a number of the training classes in Datasets A and B.

Table A1 Class Mean Vectors for the Classes in Dataset A - May 1987

	B1	B2	B3	B4	B5	B7
Water 1	92.46	44.94	49.73	21.89	8.69	3.63
B. Roof	254.77	177.42	237.77	187.65	219.85	109.81
D. Veg	80.07	33.82	26.55	139.00	77.02	20.36
C. Veg	83.03	32.77	29.10	83.43	58.92	16.93
Asphalt	125.79	51.47	63.17	51.27	61.90	35.94
Concrete	193.27	103.27	142.96	106.12	170.10	105.00
Water 2	81.03	30.66	25.72	11.83	4.50	1.70
Grass-A	106.71	49.87	60.37	103.85	124.46	49.71
Grass-B	86.42	39.75	32.62	129.71	97.62	31.12

Table A2 Class Mean Vectors for the Classes in Dataset B - May 1987

	B1	B2	B3	B4	B5	B7
Baresoil	110.76	67.11	104.37	89.66	130.00	56.84
Fields-A	106.11	51.17	67.46	98.15	108.23	53.14
Fields-C	120.49	62.86	93.33	86.57	141.27	81.49
Fields-D	88.03	39.21	34.09	154.96	98.73	29.34
Grass-A	106.71	49.87	60.37	103.85	124.46	49.71
Grass-B	86.42	39.75	32.63	129.71	97.63	31.13
Grass-C	89.61	40.77	36.97	140.77	91.39	29.32
Leaf	79.24	33.84	26.67	130.97	76.51	20.91
Pine	81.88	32.20	29.05	82.20	61.92	20.35
Road-A	122.92	54.58	68.63	66.94	88.58	49.93
Runway-C	110.01	42.47	48.74	35.08	52.15	34.13
Runway-F	173.62	93.31	133.32	113.26	176.60	100.60
Swamp-A	82.04	30.49	31.81	43.94	47.87	19.76
Swamp-B	91.67	41.08	38.25	86.58	74.67	27.92
Urban-D	220.59	114.15	150.82	110.67	177.52	122.59
Urban-F	176.93	81.47	105.93	85.33	113.93	53.80
Urban-I	185.57	95.43	130.21	102.57	152.86	90.86
Water-A1	84.57	31.38	28.31	14.45	9.41	5.02
Water-A2	80.27	29.21	25.98	12.95	6.06	2.96
Water-C	140.23	71.00	70.62	21.54	9.15	4.54

Table A3 Class Mean Vectors for the Classes in Dataset B - May 1985

	B1	B2	B3	B4	B5	B7
Baresoil	127.24	73.16	112.53	101.13	162.40	62.03
Fields-A	125.65	59.11	76.15	87.54	128.30	65.93
Fields-C	103.63	42.87	38.79	182.15	101.66	28.64
Fields-D	109.78	44.76	41.87	159.91	103.34	30.56
Grass-A	111.81	46.60	51.75	88.49	121.58	48.83
Grass-B	97.54	42.92	36.38	145.06	100.75	30.96
Grass-C	114.65	52.32	58.39	110.94	150.16	61.03
Leaf	92.28	35.09	29.46	139.97	81.23	21.06
Pine	93.92	35.14	31.65	91.70	63.09	19.62
Road-A	129.76	54.65	65.01	68.87	87.51	46.75
Runway-C	126.04	48.49	54.86	39.62	52.12	31.56
Runway-F	180.42	94.74	136.23	114.43	182.90	104.94
Swamp-A	93.56	33.06	31.74	46.61	31.72	11.47
Swamp-B	99.08	40.58	35.50	96.92	74.83	24.17
Urban-D	228.00	116.44	155.04	115.15	165.76	128.96
Urban-F	219.33	103.60	136.20	109.40	150.40	73.80
Urban-I	249.43	128.36	170.07	120.57	179.79	106.93
Water-A1	103.27	37.91	34.86	19.00	13.72	7.47
Water-A2	106.26	38.83	37.74	20.31	12.21	6.36
Water-C	149.00	72.39	78.23	28.08	15.23	7.39

Table A4 Class Mean Vectors for the Classes in Dataset B - Aug 1985

	B1	B2	B3	B4	B5	B7
Baresoil	150.34	71.95	101.76	107.26	186.26	101.84
Fields-A	145.41	53.77	51.63	141.79	95.52	26.90
Fields-C	140.60	53.25	50.61	125.35	98.77	29.90
Fields-D	144.30	53.65	52.36	120.02	106.85	32.73
Grass-A	146.76	57.92	65.87	85.03	128.72	50.83
Grass-B	144.00	55.79	55.96	107.00	83.79	26.08
Grass-C	129.84	49.94	55.07	86.74	135.71	55.26
Leaf	131.60	47.00	44.03	106.16	71.36	18.62
Pine	126.52	45.50	43.15	82.77	53.47	15.24
Road-A	147.15	57.16	64.82	66.92	81.60	41.66
Runway-C	147.04	54.28	58.45	43.68	44.81	26.06
Runway-F	171.39	78.42	106.50	96.94	161.31	90.23
Swamp-A	134.33	48.95	49.21	62.64	45.60	15.55
Swamp-B	126.00	45.25	43.25	85.50	69.75	21.67
Urban-D	185.93	82.85	106.78	96.44	158.19	107.11
Urban-F	178.20	76.73	97.93	87.67	134.40	65.20
Urban-I	188.93	88.07	117.36	109.43	162.71	93.50
Water-A1	145.37	52.90	51.67	28.94	13.37	4.76
Water-A2	148.78	53.73	53.62	32.95	14.20	5.04
Water-C	145.69	56.39	52.54	41.77	20.54	7.92

Table A5 Class Mean Vectors for the Classes in Dataset B - Oct 1985

	B1	B2	B3	B4	B5	B7
Baresoil	87.79	51.34	80.32	72.18	127.05	66.55
Fields-A	74.85	31.13	42.35	46.23	90.91	44.32
Fields-C	66.51	27.64	26.98	77.71	74.20	24.90
Fields-D	67.15	27.74	25.66	97.73	82.02	25.45
Grass-A	73.20	30.00	31.58	77.10	84.67	29.70
Grass-B	64.33	26.17	27.54	58.63	59.42	20.63
Grass-C	73.26	30.97	34.65	71.61	93.74	34.10
Leaf	61.80	23.20	23.29	65.31	51.38	14.62
Pine	61.46	22.06	20.53	49.73	34.04	10.87
Road-A	85.82	34.94	40.82	42.55	54.31	28.11
Runway-C	74.68	26.42	28.06	19.92	28.00	18.33
Runway-F	131.39	69.39	98.19	82.87	131.04	73.29
Swamp-A	65.82	23.40	25.63	30.58	39.58	15.25
Swamp-B	65.67	24.00	23.17	34.25	40.42	16.58
Urban-D	153.33	78.22	102.11	76.15	124.89	84.48
Urban-F	136.67	64.13	81.87	67.27	93.67	44.53
Urban-I	109.14	55.36	76.00	59.93	93.36	50.79
Water-A1	65.17	22.74	19.55	8.57	3.85	1.62
Water-A2	60.08	20.70	18.64	8.42	3.68	1.46
Water-C	100.92	47.54	50.62	17.39	7.85	3.62

Table A6 Class Mean Vectors for the Classes in Dataset B - March 1989

	B1	B2	B3	B4	B5	B7
Baresoil	103.16	57.11	88.61	75.29	141.63	76.40
Fields-A	100.17	45.20	57.28	70.91	107.79	48.35
Fields-C	110.20	48.03	63.89	62.84	107.06	48.77
Fields-D	126.27	54.38	71.30	84.31	123.81	51.85
Grass-A	100.53	43.79	54.45	64.66	105.61	44.89
Grass-B	94.21	38.88	48.75	52.46	97.67	42.79
Grass-C	100.55	44.19	58.74	66.61	130.42	56.58
Leaf	95.93	37.76	46.40	51.30	87.54	37.31
Pine	89.91	35.28	36.86	55.41	51.97	19.62
Road-A	105.42	44.31	54.62	45.75	68.76	37.68
Runway-C	101.91	40.39	46.03	32.90	44.58	26.97
Runway-F	137.33	66.96	93.34	75.89	124.67	67.23
Swamp-A	86.00	32.60	35.78	29.76	41.04	17.51
Swamp-B	82.33	29.92	29.17	24.17	21.83	10.17
Urban-D	156.96	76.26	103.74	76.82	129.74	82.41
Urban-F	147.73	67.73	89.33	71.33	115.13	55.33
Urban-I	135.57	69.07	96.43	74.29	120.93	68.93
Water-A1	86.63	33.37	30.20	15.89	6.93	3.42
Water-A2	86.49	33.63	31.03	15.38	5.37	2.38
Water-C	115.85	54.15	58.85	21.23	9.92	4.23

Table A7 Covariance Matrices for Classes in Dataset A - May 1987

Water 1

	B1	B2	B3	B4	B5	B7
B1	**4.23**	1.73	1.53	1.92	3.58	2.01
B2	1.73	**1.35**	0.92	0.78	1.68	1.08
B3	1.53	0.92	**1.69**	-0.38	0.44	0.43
B4	1.92	0.78	-0.38	**9.07**	9.59	4.63
B5	3.58	1.68	0.44	9.59	**14.60**	6.67
B7	2.01	1.08	0.43	4.63	6.67	**4.38**

B. Roof

	B1	B2	B3	B4	B5	B7
B1	**0.42**	-1.86	-3.70	-2.76	-4.88	-2.13
B2	-1.86	**424.09**	508.94	450.23	466.51	217.16
B3	-3.70	508.94	**643.30**	549.00	563.84	254.03
B4	-2.76	450.23	549.00	**486.40**	504.66	230.85
B5	-4.88	466.51	563.84	504.66	**826.54**	369.17
B7	-2.13	217.16	254.03	230.85	369.17	**174.88**

D. Veg

	B1	B2	B3	B4	B5	B7
B1	**1.42**	0.37	0.23	2.15	0.19	0.13
B2	0.37	**0.69**	0.32	0.86	0.65	0.22
B3	0.23	0.32	**0.62**	0.10	0.60	0.33
B4	2.15	0.86	0.10	**29.69**	7.97	0.27
B5	0.19	0.65	0.60	7.97	**8.86**	1.47
B7	0.13	0.22	0.33	0.27	1.47	**1.12**

C. Veg

	B1	B2	B3	B4	B5	B7
B1	**1.93**	-0.06	0.27	-1.66	-0.42	0.21
B2	-0.06	**0.62**	-0.01	0.51	0.34	0.14
B3	0.27	-0.01	**0.80**	-1.45	-0.38	0.14
B4	-1.66	0.51	-1.45	**17.27**	5.70	0.74
B5	-0.42	0.34	-0.38	5.70	**8.72**	2.03
B7	0.21	0.14	0.14	0.74	2.03	**1.49**

Covariance Matrices for Classes in Dataset A (continued)

Asphalt

	B1	B2	B3	B4	B5	B7
B1	**24.87**	10.71	16.05	2.23	8.62	5.12
B2	10.71	**6.60**	8.83	3.41	5.63	3.00
B3	16.05	8.83	**14.06**	4.08	8.05	4.65
B4	2.23	3.41	4.08	**13.51**	8.82	2.78
B5	8.62	5.63	8.05	8.82	**13.05**	4.81
B7	5.12	3.00	4.65	2.78	4.81	**4.14**

Concrete

	B1	B2	B3	B4	B5	B7
B1	**39.68**	16.83	16.92	3.43	6.37	4.00
B2	16.83	**11.80**	14.64	3.84	11.20	7.61
B3	16.92	14.64	**23.67**	6.37	21.80	16.31
B4	3.43	3.84	6.37	**4.91**	8.40	4.94
B5	6.37	11.20	21.80	8.40	**34.82**	23.36
B7	4.00	7.61	16.31	4.94	23.36	**21.30**

Water 2

	B1	B2	B3	B4	B5	B7
B1	**2.03**	-0.29	0.26	0.02	0.27	0.21
B2	-0.29	**0.39**	0.06	-0.14	-0.19	-0.13
B3	0.26	0.06	**0.48**	-0.15	-0.01	0.03
B4	0.02	-0.14	-0.15	**0.68**	0.10	0.07
B5	0.27	-0.19	-0.01	0.10	**0.97**	0.00
B7	0.21	-0.13	0.03	0.07	0.00	**0.69**

Grass - A

	B1	B2	B3	B4	B5	B7
B1	**23.30**	13.32	25.73	17.55	32.76	16.67
B2	13.32	**9.72**	17.38	14.93	20.85	9.22
B3	25.73	17.38	**35.89**	29.08	38.58	17.72
B4	17.55	14.93	29.08	**69.41**	17.57	-7.49
B5	32.76	20.85	38.58	17.57	**75.51**	39.33
B7	16.67	9.22	17.72	-7.49	39.33	**28.88**

Table A8 Correlation Matrices for Classes in Dataset A - May 1987

Water 1

	B1	B2	B3	B4	B5	B7
B1	1.00	0.73	0.57	0.31	0.46	0.47
B2	0.73	1.00	0.61	0.22	0.38	0.44
B3	0.57	0.61	1.00	-0.10	0.09	0.16
B4	0.31	0.22	-0.10	1.00	0.83	0.73
B5	0.46	0.38	0.09	0.83	1.00	0.83
B7	0.47	0.44	0.16	0.73	0.83	1.00

B. Roof

	B1	B2	B3	B4	B5	B7
B1	1.00	-0.14	-0.22	-0.19	-0.26	-0.25
B2	-0.14	1.00	0.97	0.99	0.79	0.80
B3	-0.22	0.97	1.00	0.98	0.77	0.76
B4	-0.19	0.99	0.98	1.00	0.80	0.79
B5	-0.26	0.79	0.77	0.80	1.00	0.97
B7	-0.25	0.80	0.76	0.79	0.97	1.00

D. Veg

	B1	B2	B3	B4	B5	B7
B1	1.00	0.37	0.25	0.33	0.05	0.10
B2	0.37	1.00	0.49	0.19	0.26	0.25
B3	0.25	0.49	1.00	0.02	0.26	0.39
B4	0.33	0.19	0.02	1.00	0.49	0.05
B5	0.05	0.26	0.26	0.49	1.00	0.47
B7	0.10	0.25	0.39	0.05	0.47	1.00

C. Veg

	B1	B2	B3	B4	B5	B7
B1	1.00	-0.05	0.22	-0.29	-0.10	0.12
B2	-0.05	1.00	-0.01	0.16	0.14	0.14
B3	0.22	-0.01	1.00	-0.39	-0.14	0.13
B4	-0.29	0.16	-0.39	1.00	0.46	0.15
B5	-0.10	0.14	-0.14	0.46	1.00	0.56
B7	0.12	0.14	0.13	0.15	0.56	1.00

Correlation Matrices for Classes in Bands A (continued)

Asphalt

	B1	B2	B3	B4	B5	B7
B1	1.00	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
B2	[illegible]	1.00	0.92	[illegible]	[illegible]	[illegible]
B3	[illegible]	0.92	1.00	[illegible]	[illegible]	[illegible]
B4	0.12	[illegible]	[illegible]	1.00	[illegible]	[illegible]
B5	[illegible]	[illegible]	[illegible]	[illegible]	1.00	[illegible]
B7	[illegible]	0.57	[illegible]	[illegible]	[illegible]	1.00

Concrete

	B1	B2	B3	B4	B5	B7
B1	1.00	0.76	[illegible]	[illegible]	[illegible]	[illegible]
B2	0.76	1.00	[illegible]	[illegible]	[illegible]	[illegible]
B3	0.55	[illegible]	1.00	[illegible]	[illegible]	[illegible]
B4	0.25	[illegible]	[illegible]	1.00	[illegible]	[illegible]
B5	0.17	[illegible]	[illegible]	[illegible]	1.00	[illegible]
B7	0.14	[illegible]	[illegible]	[illegible]	[illegible]	1.00

Water 1

	B1	B2	B3	B4	B5	B7
B1	1.00	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
B2	[illegible]	1.00	[illegible]	[illegible]	[illegible]	[illegible]
B3	[illegible]	[illegible]	1.00	[illegible]	[illegible]	[illegible]
B4	[illegible]	[illegible]	[illegible]	1.00	[illegible]	[illegible]
B5	[illegible]	[illegible]	[illegible]	[illegible]	1.00	[illegible]
B7	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	1.00

GRASS-A

	B1	B2	B3	B4	B5	B7
B1	1.00	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
B2	[illegible]	1.00	[illegible]	[illegible]	[illegible]	[illegible]
B3	[illegible]	[illegible]	1.00	[illegible]	[illegible]	[illegible]
B4	[illegible]	[illegible]	[illegible]	1.00	[illegible]	[illegible]
B5	[illegible]	[illegible]	[illegible]	[illegible]	1.00	[illegible]
B7	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	1.00

APPENDIX D. Supporting Data for [illegible]

[illegible]

Table D4. Contingency Table Results for [illegible]

BAYES CONTINGENCY RESULTS [illegible]

CLASS	Water 1	B. Sand	B. Veg	C. Veg	Asphalt	Concrete	Water 2	Grass A	TOTAL
Water 1	100	0	0	0	0	0	0	0	100
B. Sand	0	[illegible]	0	0	0	0	0	0	[illegible]
B. Veg	0	0	[illegible]	0	0	0	0	0	[illegible]
C. Veg	0	0	0	[illegible]	0	0	0	0	[illegible]
Asphalt	0	0	0	0	[illegible]	0	0	0	[illegible]
Concrete	0	0	0	0	0	[illegible]	0	0	[illegible]
Water 2	0	0	0	0	0	0	[illegible]	0	[illegible]
Grass A	0	0	0	0	0	0	0	[illegible]	[illegible]

MAHALANOBIS CONTINGENCY RESULTS [illegible]

CLASS	Water 1	B. Sand	B. Veg	C. Veg	Asphalt	Concrete	Water 2	Grass A	TOTAL
Water 1	100	0	0	0	0	0	0	0	[illegible]
B. Sand	0	[illegible]	0	0	0	0	0	0	[illegible]
B. Veg	0	0	[illegible]	0	0	0	0	0	[illegible]
C. Veg	0	0	0	[illegible]	0	0	0	0	[illegible]
Asphalt	0	0	0	0	[illegible]	0	0	0	[illegible]
Concrete	0	0	0	0	0	[illegible]	0	0	[illegible]
Water 2	0	0	0	0	0	0	[illegible]	0	[illegible]
Grass A	0	0	0	0	0	0	0	[illegible]	[illegible]

EUCLIDEAN CONTINGENCY RESULTS [illegible]

CLASS	Water 1	B. Sand	B. Veg	C. Veg	Asphalt	Concrete	Water 2	Grass A	TOTAL
Water 1	100	0	0	0	0	0	0	0	[illegible]
B. Sand	0	[illegible]	0	0	0		0	0	[illegible]
B. Veg	0	0	[illegible]	0	0	0	0	0	[illegible]
C. Veg	0	0	0	[illegible]	0	0	0	0	[illegible]
Asphalt	0	0	0	0	[illegible]	0	0	0	[illegible]
Concrete	0	0	0	0	0	[illegible]	0	0	[illegible]
Water 2	0	0	0	0	0	0	[illegible]	0	[illegible]
Grass A	0	0	0	0	0	0	0	[illegible]	[illegible]

Contingency Table Results for Trial #1 (continued)

Table B2 - *ii*
MAHALANOBIS CONTINGENCY RESULTS - Test Data = Set B2

	Water 1	B. Roof	D. Veg	C. Veg	Asphalt	Concete	Water 2	TOTAL
Construction	0	0	0	0	34	0	0	34
TEC Site	0	0	0	0	0	26	0	26
Parkland 1	0	0	60	0	0	0	0	60
High School	0	0	0	0	4	24	0	28
Mall	0	0	0	0	33	29	0	62
Parkland 2	0	0	69	0	0	0	0	69
Baresoil	0	0	0	0	0	38	0	38
Fields-A	0	0	260	1	146	593	0	1000
Fields-C	0	0	0	0	48	53	0	101
Fields-D	0	0	105	0	1	0	0	106
Grass-A	0	0	0	0	35	52	0	87
Grass-B	0	0	23	0	1	0	0	24
Grass-C	0	0	29	0	2	0	0	31
Leaf	0	0	958	42	0	0	0	1000
Pine	0	0	2	387	4	0	0	393
Road-A	0	0	1	2	432	66	0	501
Runway-C	0	0	0	0	78	0	0	78
Runway-F	0	0	0	0	0	97	0	97
Swamp-A	45	0	0	95	527	0	4	671
Swamp-B	0	0	0	4	8	0	0	12
Urban-D	0	0	0	0	0	27	0	27
Urban-F	0	0	0	0	7	8	0	15
Urban-I	0	0	0	0	0	14	0	14
Water-A1	19	0	0	0	182	0	799	1000
Water-A2	0	0	0	0	27	0	973	1000
Water-C	11	0	0	0	2	0	0	13

Contingency Table Results for Trial #1 (continued)

Table B2 - *iii*
EUCLIDEAN CONTINGENCY RESULTS - Test Data = Set B2

	Water 1	B. Roof	D. Veg	C. Veg	Asphalt	Concete	Water 2	TOTAL
Construction	0	0	0	18	16	0	0	34
TEC Site	0	0	0	0	13	13	0	26
Parkland 1	0	0	59	1	0	0	0	60
High School	0	0	0	0	15	13	0	28
Mall	0	0	0	0	44	18	0	62
Parkland 2	0	0	69	0	0	0	0	69
BareSoil	0	0	0	0	37	1	0	38
Fields-A	0	0	344	23	532	101	0	1000
Fields-C	0	0	0	0	49	52	0	101
Fields-D	0	0	106	0	0	0	0	106
Grass-A	0	0	67	4	16	0	0	87
Grass-B	0	0	24	0	0	0	0	24
Grass-C	0	0	31	0	0	0	0	31
Leaf	0	0	945	55	0	0	0	1000
Pine	0	0	0	393	0	0	0	393
Road-A	0	0	5	13	475	8	0	501
Runway-C	0	0	0	0	78	0	0	78
Runway-F	0	0	0	0	0	97	0	97
Swamp-A	67	0	0	360	12	0	232	671
Swamp-B	0	0	0	12	0	0	0	12
Urban-D	0	0	0	0	0	27	0	27
Urban-F	0	0	0	0	7	8	0	15
Urban-I	0	0	0	0	0	14	0	14
Water-A1	113	0	0	0	0	0	887	1000
Water-A2	0	0	0	0	0	0	1000	1000
Water-C	13	0	0	0	0	0	0	13

Table B3 Contingency Table Results for Trial #2

Table B3 - *i*
BAYES CONTINGENCY RESULTS - Test Data = Set B2

CLASS	Water 1	B.Roof	D. Veg	C. Veg	Asphalt	Concrete	Water 2	Grass-A	TOTAL
Construction	0	0	0	0	25	0	0	9	34
TEC Site	0	0	0	0	0	4	0	22	26
Parkland 1	0	0	60	0	0	0	0	0	60
High School	0	0	0	0	3	22	0	3	28
Mall	0	0	0	0	30	23	0	9	62
Parkland 2	0	0	1	0	0	0	0	68	69
BareSoil	0	0	0	0	0	33	0	5	38
Fields-A	0	0	260	1	0	104	0	635	1000
Fields-C	0	0	0	0	0	50	0	51	101
Fields-D	0	0	76	0	0	0	0	30	106
Grass-B	0	0	7	0	0	0	0	17	24
Grass-C	0	0	18	0	0	0	0	13	31
Leaf	0	0	928	14	0	0	0	58	1000
Pine	0	0	2	346	0	0	0	45	393
Road-A	0	0	0	0	325	17	0	159	501
Runway-C	0	0	0	0	78	0	0	0	78
Runway-F	0	0	0	0	0	93	0	4	97
Swamp-A	45	0	0	77	309	0	4	236	671
Swamp-B	0	0	0	1	0	0	0	11	12
Urban-D	0	0	0	0	0	27	0	0	27
Urban-F	0	0	0	0	7	8	0	0	15
Urban-I	0	0	0	0	0	14	0	0	14
Water-A1	19	0	0	0	169	0	812	0	1000
Water-A2	0	0	0	0	26	0	974	0	1000
Water-C	11	0	0	0	2	0	0	0	17

Table B3 [illegible]

[illegible]

CLASS	Water [illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Construction	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	
[illegible] Site	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	
Parkland [illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	
High School	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	
Mall	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	
Parkland [illegible]	[illegible]	[illegible]		[illegible]	[illegible]		
Baseball	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]		
Fields A	[illegible]	[illegible]	[illegible]			[illegible]	
Fields [illegible]	[illegible]	[illegible]	[illegible]	[illegible]		[illegible]	
Fields [illegible]	[illegible]	[illegible]	[illegible]	[illegible]			
Grass [illegible]	[illegible]	[illegible]		[illegible]			
Grass [illegible]	[illegible]	[illegible]	[illegible]	[illegible]			
Leaf	[illegible]	[illegible]	[illegible]				
Pine	[illegible]	[illegible]		[illegible]			
Road [illegible]	[illegible]	[illegible]	[illegible]	[illegible]			
Runway [illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]		
Runway [illegible]	[illegible]	[illegible]	[illegible]	[illegible]			
Swamp [illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]		
Swamp [illegible]	[illegible]	[illegible]	[illegible]				
Urban [illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]		
Urban [illegible]	[illegible]	[illegible]	[illegible]				
Urban [illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]		
Water [illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]		
Water [illegible]	[illegible]	[illegible]	[illegible]	[illegible]			
Water [illegible]	[illegible]	[illegible]	[illegible]	[illegible]			

AD-A274 142 MULTIVARIATE SPECTRAL ANALYSIS TO EXTRACT MATERIALS FROM MULTISPECTRAL DATA(U) ARMY TOPOGRAPHIC ENGINEERING CENTER FORT BELVOIR VA R S RAND ET AL. SEP 93 TEC-0039 2/2

UNCLASSIFIED XA-TEC** NL

END
FILMED
DTIC

AIIM

Association for Information and Image Management

1100 Wayne Avenue, Suite 1100
Silver Spring, Maryland 20910

301/587-8202

Centimeter

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 mm

1 2 3 4 5

Inches

1.0 2.8 2.5 4.5 5.0 3.2 2.2 5.6 6.3 3.6 7.1 2.0 8.0 4.0 9.0 1.1 10 11 12 1.8 1.25 1.4 1.6

MANUFACTURED TO AIIM STANDARDS
BY APPLIED IMAGE, INC.

Contingency Table Results for Trial #2 (continued)

Table B3 - *iii*
EUCLIDEAN CONTINGENCY RESULTS - Test Data = Set B2

CLASS	Water 1	B.Roof	D. Veg	C. Veg	Asphalt	Concrete	Water 2	Grass-A	TOTAL
Construction	0	0	0	18	16	0	0	0	34
TEC Site	0	0	0	0	0	9	0	17	26
Parkland 1	0	0	59	1	0	0	0	0	60
High School	0	0	0	0	2	6	0	20	28
Mall	0	0	0	0	29	13	0	20	62
Parkland 2	0	0	57	0	0	0	0	12	69
BareSoil	0	0	0	0	0	0	0	38	38
Fields-A	0	0	265	1	11	61	0	662	1000
Fields-C	0	0	0	0	36	50	0	15	101
Fields-D	0	0	104	0	0	0	0	2	106
Grass-B	0	0	23	0	0	0	0	1	24
Grass-C	0	0	28	0	0	0	0	3	31
Leaf	0	0	945	55	0	0	0	0	1000
Pine	0	0	0	392	0	0	0	1	393
Road-A	0	0	1	7	354	0	0	139	501
Runway-C	0	0	0	0	78	0	0	0	78
Runway-F	0	0	0	0	0	96	0	1	97
Swamp-A	67	0	0	360	11	0	232	1	671
Swamp-B	0	0	0	12	0	0	0	0	12
Urban-D	0	0	0	0	0	27	0	0	27
Urban-F	0	0	0	0	6	8	0	1	15
Urban-I	0	0	0	0	0	14	0	0	14
Water-A1	113	0	0	0	0	0	887	0	1000
Water-A2	0	0	0	0	0	0	1000	0	1000
Water-C	13	0	0	0	0	0	0	0	13

Table B4 Contingency Table Results for Trial #3

Table B4 - *i*

MODIFIED BAYES CONTINGENCY RESULTS - Test Data = B2 with $\chi^2(6)$ =16.81
MinVar = 16 on Water; MinVar=3 on other classes

CLASS	Water 1	B.Roof	D. Veg	C. Veg	Asphalt	Concrete	Water 2	Grass-B	NULL
Construction	0	0	0	0	0	0	0	0	34
TEC Site	0	0	0	0	0	0	0	0	26
Parkland 1	0	0	33	0	0	0	0	0	27
High School	0	0	0	0	0	0	0	0	28
Mall	0	0	0	0	0	0	0	0	62
Parkland 2	0	0	0	0	0	0	0	0	69
BareSoil	0	0	0	0	0	0	0	0	38
Fields-A	0	0	9	0	0	0	0	0	991
Fields-C	0	0	0	0	0	0	0	0	101
Fields-D	0	0	0	0	0	0	0	34	72
Grass-A	0	0	0	0	0	0	0	0	87
Grass-C	0	0	0	0	0	0	0	6	25
Leaf	0	0	712	0	0	0	0	39	249
Pine	0	0	0	262	0	0	0	0	131
Road-A	0	0	0	0	67	0	0	0	434
Runway-C	0	0	0	0	0	0	0	0	78
Runway-F	0	0	0	0	0	2	0	0	95
Swamp-A	0	0	0	0	0	0	6	0	665
Swamp-B	0	0	0	0	0	0	0	0	12
Urban-D	0	0	0	0	0	0	0	0	27
Urban-F	0	0	0	0	0	0	0	0	15
Urban-I	0	0	0	0	0	3	0	0	11
Water-A1	4	0	0	0	0	0	870	0	126
Water-A2	0	0	0	0	0	0	993	0	7
Water-C	0	0	0	0	0	0	0	0	13
TOTAL	4	0	754	262	67	5	1869	79	3423

Contingency Table Results for Trial #3 (continued)

Table B4 - *ii*

MODIFIED BAYES CONTINGENCY RESULTS - Test Data = B2 with $\chi^2(6) = 84.05$

MinVar = 16 on Water; MinVar=3 on other classes

CLASS	Water 1	B.Roof	D. Veg	C. Veg	Asphalt	Concrete	Water 2	Grass-B	NULL
Construction	0	0	0	0	20	0	0	0	14
TEC Site	0	0	0	0	0	0	0	0	26
Parkland 1	0	0	60	0	0	0	0	0	0
High School	0	0	0	0	0	1	0	0	27
Mall	0	0	0	0	1	1	0	0	60
Parkland 2	0	0	0	0	0	0	0	59	10
BareSoil	0	0	0	0	0	0	0	0	38
Fields-A	0	0	261	0	0	3	0	78	658
Fields-C	0	0	0	0	0	0	0	0	101
Fields-D	0	0	2	0	0	0	0	104	0
Grass-A	0	0	0	0	0	0	0	34	53
Grass-C	0	0	1	0	0	0	0	29	1
Leaf	0	0	844	23	0	0	0	133	0
Pine	0	0	2	382	0	0	0	4	5
Road-A	0	0	0	0	164	0	0	9	328
Runway-C	0	0	0	0	72	0	0	0	6
Runway-F	0	0	0	0	0	75	0	0	22
Swamp-A	74	0	0	51	0	0	163	0	383
Swamp-B	0	0	0	5	0	0	0	1	6
Urban-D	0	0	0	0	0	10	0	0	17
Urban-F	0	0	0	0	0	1	0	0	14
Urban-I	0	0	0	0	0	13	0	0	1
Water-A1	110	0	0	0	0	0	890	0	0
Water-A2	0	0	0	0	0	0	1000	0	0
Water-C	0	0	0	0	0	0	0	0	13
TOTAL	184	0	1170	461	257	104	2053	451	1783

Contingency Table Results for Trial #3 (continued)

Table B4 - *iii*

MODIFIED BAYES CONTINGENCY RESULTS - Test Data = B2 with $\chi^2(6) = 117.67$
MinVar = 16 on Water; MinVar=3 on other classes

CLASS	Water 1	B.Roof	D. Veg	C. Veg	Asphalt	Concrete	Water 2	Grass-B	NULL
Construction	0	0	0	0	28	0	0	0	6
TEC Site	0	0	0	0	0	0	0	0	26
Parkland 1	0	0	60	0	0	0	0	0	0
High School	0	0	0	0	0	4	0	0	24
Mall	0	0	0	0	3	5	0	0	54
Parkland 2	0	0	0	0	0	0	0	68	1
BareSoil	0	0	0	0	0	0	0	0	38
Fields-A	0	0	261	0	0	50	0	112	577
Fields-C	0	0	0	0	0	8	0	0	93
Fields-D	0	0	2	0	0	0	0	104	0
Grass-A	0	0	0	0	0	0	0	53	34
Grass-C	0	0	1	0	0	0	0	29	1
Leaf	0	0	844	23	0	0	0	133	0
Pine	0	0	2	383	0	0	0	7	1
Road-A	0	0	0	0	217	0	0	13	271
Runway-C	0	0	0	0	77	0	0	0	1
Runway-F	0	0	0	0	0	84	0	0	13
Swamp-A	84	0	0	107	16	0	215	1	248
Swamp-B	0	0	0	6	0	0	0	5	1
Urban-D	0	0	0	0	0	26	0	0	1
Urban-F	0	0	0	0	0	2	0	0	13
Urban-I	0	0	0	0	0	14	0	0	0
Water-A1	110	0	0	0	0	0	890	0	0
Water-A2	0	0	0	0	0	0	1000	0	0
Water-C	0	0	0	0	0	0	0	0	13
TOTAL	**194**	**0**	**1170**	**519**	**341**	**193**	**2105**	**525**	**1416**

Contingency Table Results for Trial #3 (continued)

Table B4 - *iv*

MODIFIED BAYES CONTINGENCY RESULTS - Test Data =Set B2 ; with $\chi^2(6) = \infty$

MinVar=16 on Water; MinVar = 3 on other classes

CLASS	Water 1	B.Roof	D. Veg	C. Veg	Asphalt	Concrete	Water 2	Grass-B	TOTAL
Construction	0	0	0	0	33	0	0	1	34
TEC Site	0	0	0	0	0	25	0	1	26
Parkland 1	0	0	60	0	0	0	0	0	60
High School	0	0	0	0	4	24	0	0	28
Mall	0	0	0	0	33	28	0	1	62
Parkland 2	0	0	0	0	0	0	0	69	69
BareSoil	0	0	0	0	0	38	0	0	38
Fields-A	0	0	261	0	20	412	0	307	1000
Fields-C	0	0	0	0	17	52	0	32	101
Fields-D	0	0	2	0	0	0	0	104	106
Grass-A	0	0	0	0	0	0	0	87	87
Grass-C	0	0	1	0	0	0	0	30	31
Leaf	0	0	844	23	0	0	0	133	1000
Pine	0	0	2	383	0	0	0	8	393
Road-A	0	0	0	0	407	59	0	35	501
Runway-C	0	0	0	0	78	0	0	0	78
Runway-F	0	0	0	0	0	97	0	0	97
Swamp-A	88	0	0	174	102	0	229	78	671
Swamp-B	0	0	0	6	0	0	0	6	12
Urban-D	0	0	0	0	0	27	0	0	27
Urban-F	0	0	0	0	7	8	0	0	15
Urban-I	0	0	0	0	0	14	0	0	14
Water-A1	110	0	0	0	0	0	890	0	1000
Water-A2	0	0	0	0	0	0	1000	0	1000
Water-C	13	0	0	0	0	0	0	0	13
TOTAL	211	0	1170	586	701	784	2119	892	6463

Contingency Table Results for Trial #3 (continued)

Table B4 -v
STANDARD BAYES CONTINGENCY RESULTS - Test Data = Set B2 (No minimum variance or rejection criteria)

CLASS	Water 1	B.Roof	D. Veg	C. Veg	Asphalt	Concrete	Water 2	Grass-B	TOTAL
Construction	0	0	0	0	33	0	0	1	34
TEC Site	0	0	0	0	0	25	0	1	26
Parkland 1	0	0	60	0	0	0	0	0	60
High School	0	0	0	0	4	24	0	0	28
Mall	0	0	0	0	33	28	0	1	62
Parkland 2	0	0	0	0	0	0	0	69	69
BareSoil	0	0	0	0	0	38	0	0	38
Fields-A	0	0	241	1	21	418	0	319	1000
Fields-C	0	0	0	0	17	52	0	32	101
Fields-D	0	0	1	0	0	0	0	105	106
Grass-A	0	0	0	0	0	0	0	87	87
Grass-C	0	0	0	0	0	0	0	31	31
Leaf	0	0	806	22	0	0	0	172	1000
Pine	0	0	2	378	0	0	0	13	393
Road-A	0	0	0	0	408	60	0	33	501
Runway-C	0	0	0	0	78	0	0	0	78
Runway-F	0	0	0	0	0	97	0	0	97
Swamp-A	45	0	0	108	446	0	4	68	671
Swamp-B	0	0	0	3	0	0	0	9	12
Urban-D	0	0	0	0	0	27	0	0	27
Urban-F	0	0	0	0	7	8	0	0	15
Urban-I	0	0	0	0	0	14	0	0	14
Water-A1	19	0	0	0	169	0	812	0	1000
Water-A2	0	0	0	0	26	0	974	0	1000
Water-C	11	0	0	0	2	0	0	0	13
TOTAL	75	0	1110	512	1244	791	1790	941	6463

Contingency Table Results for Trial #3 (continued)

Table B4 - *vi*
MAHALANOBIS CONTINGENCY RESULTS - Test Data = Set B2

CLASS	Water 1	B.Roof	D. Veg	C. Veg	Asphalt	Concrete	Water 2	Grass-B	TOTAL
Construction	0	0	0	0	33	0	0	1	34
TEC Site	0	0	0	0	0	25	0	1	26
Parkland 1	0	0	60	0	0	0	0	0	60
High School	0	0	0	0	4	24	0	0	28
Mall	0	0	0	0	33	28	0	1	62
Parkland 2	0	0	0	0	0	0	0	69	69
BareSoil	0	0	0	0	0	38	0	0	38
Fields-A	0	0	222	1	20	420	0	337	1000
Fields-C	0	0	0	0	16	52	0	33	101
Fields-D	0	0	0	0	0	0	0	106	106
Grass-A	0	0	0	0	0	0	0	87	87
Grass-C	0	0	0	0	0	0	0	31	31
Leaf	0	0	766	19	0	0	0	215	1000
Pine	0	0	2	376	0	0	0	15	393
Road-A	0	0	0	0	408	60	0	33	501
Runway-C	0	0	0	0	78	0	0	0	78
Runway-F	0	0	0	0	0	97	0	0	97
Swamp-A	45	0	0	94	460	0	4	68	671
Swamp-B	0	0	0	3	0	0	0	9	12
Urban-D	0	0	0	0	0	27	0	0	27
Urban-F	0	0	0	0	7	8	0	0	15
Urban-I	0	0	0	0	0	14	0	0	14
Water-A1	19	0	0	0	182	0	799	0	1000
Water-A2	0	0	0	0	27	0	973	0	1000
Water-C	11	0	0	0	2	0	0	0	13
TOTAL	75	0	1050	493	1270	793	1776	1006	6463

Contingency Table Results for Trial #3 (continued)

Table B4 - *vii*
EUCLIDEAN CONTINGENCY RESULTS - Test Data = Set B2

CLASS	Water 1	B.Roof	D. Veg	C. Veg	Asphalt	Concrete	Water 2	Grass-B	TOTAL
Construction	0	0	0	18	16	0	0	0	34
TEC Site	0	0	0	0	9	13	0	4	26
Parkland 1	0	0	59	1	0	0	0	0	60
High School	0	0	0	0	15	13	0	0	28
Mall	0	0	0	0	42	18	0	2	62
Parkland 2	0	0	0	0	0	0	0	69	69
BareSoil	0	0	0	0	37	1	0	0	38
Fields-A	0	0	260	1	502	99	0	138	1000
Fields-C	0	0	0	0	48	52	0	1	101
Fields-D	0	0	39	0	0	0	0	67	106
Grass-A	0	0	0	0	0	0	0	87	87
Grass-C	0	0	12	0	0	0	0	19	31
Leaf	0	0	885	44	0	0	0	71	1000
Pine	0	0	0	393	0	0	0	0	393
Road-A	0	0	0	7	469	8	0	17	501
Runway-C	0	0	0	0	78	0	0	0	78
Runway-F	0	0	0	0	0	97	0	0	97
Swamp-A	67	0	0	360	12	0	232	0	671
Swamp-B	0	0	0	9	0	0	0	3	12
Urban-D	0	0	0	0	0	27	0	0	27
Urban-F	0	0	0	0	7	8	0	0	15
Urban-I	0	0	0	0	0	14	0	0	14
Water-A1	113	0	0	0	0	0	887	0	1000
Water-A2	0	0	0	0	0	0	1000	0	1000
Water-C	13	0	0	0	0	0	0	0	13
TOTAL	193	0	1255	833	1235	350	2119	478	6463

Table B5 Contingency Table Results for Trial #4

MODIFIED BAYES CONTINGENCY RESULTS - Test Data = B2 with $\chi^2(6) = 13.28$
MinVar = 16 on Water; MinVar=3 on other classes

CLASS	Water 1	B.Roof	D. Veg	C. Veg	Asphalt	Concrete	Water 2	Grass-B	NULL
Construction	0	0	0	0	0	0	0	0	34
ETL Site	0	0	0	0	0	0	0	0	26
Parkland 1	0	0	33	0	0	0	0	0	27
High School	0	0	0	0	0	0	0	0	28
Mall	0	0	0	0	0	0	0	0	62
Parkland 2	0	0	0	0	0	0	0	0	69
BareSoil	0	0	0	0	0	0	0	0	38
Fields-A	0	0	5	0	0	0	0	0	995
Fields-C	0	0	0	0	0	0	0	0	101
Fields-D	0	0	0	0	0	0	0	43	63
Grass-A	0	0	0	0	0	0	0	0	87
Grass-C	0	0	0	0	0	0	0	6	25
Leaf	0	0	699	0	0	0	0	80	221
Pine	0	0	0	290	0	0	0	0	103
Road-A	0	0	0	0	74	0	0	1	426
Runway-C	0	0	0	0	1	0	0	0	77
Runway-F	0	0	0	0	0	7	0	0	90
Swamp-A	0	0	0	0	0	0	5	0	666
Swamp-B	0	0	0	0	0	0	0	0	12
Urban-D	0	0	0	0	0	0	0	0	27
Urban-F	0	0	0	0	0	0	0	0	15
Urban-I	0	0	0	0	0	3	0	0	11
Water-A1	2	0	0	0	0	0	874	0	124
Water-A2	0	0	0	0	0	0	996	0	4
Water-C	0	0	0	0	0	0	0	0	13
TOTAL	2	0	737	290	75	10	1875	130	3344

Contingency Table Results for Trial #4 (continued)

MODIFIED BAYES CONTINGENCY RESULTS - Test Data = B2 with $\chi^2(6) = 66.4$
MinVar = 16 on Water; MinVar=3 on other classes

CLASS	Water 1	B.Roof	D. Veg	C. Veg	Asphalt	Concrete	Water 2	Grass-B	NULL
Construction	0	0	0	0	17	0	0	0	17
ETL Site	0	5	0	0	0	0	0	0	21
Parkland 1	0	0	60	0	0	0	0	0	0
High School	0	17	0	0	0	0	0	0	11
Mall	0	39	0	0	0	3	0	0	20
Parkland 2	0	0	0	0	0	0	0	60	9
BareSoil	0	28	0	0	0	0	0	0	10
Fields-A	0	282	258	0	0	5	0	70	385
Fields-C	0	2	0	0	0	1	0	0	98
Fields-D	0	0	2	0	0	0	0	104	0
Grass-A	0	0	0	0	0	0	0	26	61
Grass-C	0	0	2	0	0	0	0	28	1
Leaf	0	0	815	11	0	0	0	174	0
Pine	0	0	2	378	0	0	0	6	7
Road-A	0	160	0	0	147	0	0	8	186
Runway-C	0	7	0	0	69	0	0	0	2
Runway-F	0	36	0	0	0	50	0	0	11
Swamp-A	101	0	0	30	0	0	116	0	424
Swamp-B	0	0	0	6	0	0	0	1	5
Urban-D	0	0	0	0	0	26	0	0	1
Urban-F	0	15	0	0	0	0	0	0	0
Urban-I	0	1	0	0	0	12	0	0	1
Water-A1	94	0	0	0	0	0	906	0	0
Water-A2	1	0	0	0	0	0	999	0	0
Water-C	13	0	0	0	0	0	0	0	0
TOTAL	209	592	1139	425	233	97	2021	477	1270

APPENDIX C: Supporting Data for Trial 5

Table C1 Auto-Classification Summary for Training Set B -Unconsolidated

No classes are combined

	Training Data MY87_1000Samples			Training Data MY85_1000Samples		
	Bayes	Mahalanobis	Euclidean	Bayes	Mahalanobis	Euclidean
Baresoil	0.00%	0.00%	0.00%	0.00%	0.00%	2.63%
Fields-A	36.80%	12.30%	83.00%	9.60%	3.20%	47.10%
Fields-C	2.97%	9.90%	92.08%	1.98%	1.98%	2.97%
Fields-D	3.77%	5.66%	11.32%	0.00%	0.00%	4.72%
Grass-A	2.30%	5.75%	2.30%	1.15%	12.64%	1.15%
Grass-B	0.00%	8.33%	16.67%	0.00%	0.00%	12.50%
Grass-C	0.00%	16.13%	16.13%	0.00%	3.23%	6.45%
Leaf	3.10%	27.30%	8.20%	1.30%	1.80%	6.50%
Pine	2.80%	8.14%	10.69%	2.80%	12.72%	17.56%
Road-A	8.38%	10.98%	30.34%	3.99%	6.99%	19.36%
Runway-C	5.13%	5.13%	5.13%	1.28%	1.28%	0.00%
Runway-F	0.00%	0.00%	4.12%	0.00%	0.00%	0.00%
Swamp-A	0.45%	0.30%	30.40%	1.34%	3.13%	9.84%
Swamp-B	0.00%	0.00%	16.67%	0.00%	0.00%	8.33%
Urban-D	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Urban-F	0.00%	0.00%	6.67%	0.00%	0.00%	0.00%
Urban-I	0.00%	0.00%	0.00%	0.00%	0.00%	7.14%
Water-A1	32.00%	10.30%	34.60%	10.40%	15.50%	34.90%
Water-A2	16.50%	42.50%	20.70%	15.00%	10.80%	34.10%
Water-C	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%

Table C1 Auto-Classification Summary for Training Set B -Unconsolidated (continued).

No classes are combined

	Training Data AG85_1000Samples			Training Data OC85_1000Samples		
	Bayes	Mahalanobis	Euclidean	Bayes	Mahalanobis	Euclidean
Baresoil	0.00%	0.00%	2.63%	0.00%	0.00%	7.89%
Fields-A	5.70%	4.20%	24.10%	0.80%	0.50%	25.80%
Fields-C	3.96%	6.93%	46.53%	4.95%	4.95%	17.82%
Fields-D	8.49%	7.55%	20.75%	5.66%	4.72%	15.09%
Grass-A	0.00%	0.00%	0.00%	11.49%	11.49%	34.48%
Grass-B	0.00%	0.00%	0.00%	4.17%	4.17%	20.83%
Grass-C	0.00%	0.00%	0.00%	6.45%	9.68%	19.35%
Leaf	3.00%	2.60%	9.10%	8.40%	8.00%	24.60%
Pine	2.80%	6.62%	11.70%	2.80%	15.27%	3.82%
Road-A	1.80%	0.80%	7.98%	2.59%	1.80%	32.14%
Runway-C	0.00%	1.28%	1.28%	1.28%	5.13%	0.00%
Runway-F	0.00%	0.00%	0.00%	0.00%	0.00%	3.09%
Swamp-A	0.45%	0.75%	2.24%	7.00%	3.73%	59.76%
Swamp-B	0.00%	0.00%	0.00%	0.00%	16.67%	0.00%
Urban-D	0.00%	0.00%	0.00%	0.00%	0.00%	3.70%
Urban-F	0.00%	0.00%	6.67%	0.00%	0.00%	0.00%
Urban-I	0.00%	0.00%	14.29%	0.00%	0.00%	0.00%
Water-A1	32.50%	12.90%	56.70%	12.50%	12.90%	15.00%
Water-A2	11.00%	29.80%	24.60%	7.00%	7.30%	17.70%
Water-C	0.00%	0.00%	7.69%	0.00%	0.00%	0.00%

Table C1 Auto-Classification Summary for Training Set B -Unconsolidated (continued).

No classes are combined

	Training Data MR89_1000Samples		
	Bayes	Mahalanobis	Euclidean
Baresoil	0.00%	0.00%	0.00%
Fields-A	11.00%	2.40%	95.30%
Fields-C	10.89%	20.79%	86.14%
Fields-D	4.72%	8.49%	59.43%
Grass-A	2.30%	21.84%	18.39%
Grass-B	0.00%	50.00%	20.83%
Grass-C	6.45%	32.26%	6.45%
Leaf	11.20%	11.80%	63.00%
Pine	6.36%	3.56%	34.86%
Road-A	10.38%	6.39%	35.93%
Runway-C	2.56%	19.23%	1.28%
Runway-F	5.15%	6.19%	16.49%
Swamp-A	2.68%	3.87%	29.81%
Swamp-B	0.00%	8.33%	0.00%
Urban-D	0.00%	0.00%	7.41%
Urban-F	0.00%	0.00%	6.67%
Urban-I	0.00%	0.00%	7.14%
Water-A1	47.90%	14.10%	57.60%
Water-A2	10.10%	69.10%	18.10%
Water-C	0.00%	0.00%	0.00%

APPENDIX D: Linear Model Results for Two Endmembers

Regression and ANOVA Tables Used in the Mixture Analysis

DEP VAR: Swamp C174 N: 6 MULTIPLE R: 0.990 SQUARED MULTIPLE R: 0.981
ADJUSTED SQUARED MULTIPLE R: 0.968 STANDARD ERROR OF ESTIMATE: 5.027

VARIABLE	COEFFICIENT	STD ERROR	STD COEF	TOLERANCE	T	P(2 TAIL)
CONSTANT	5.993	4.123	0.000	.	1.454	0.242
Leaf C133	0.196	0.047	0.337	0.977	4.184	0.025
Water B190	0.710	0.065	0.881	0.977	10.934	0.002

ANALYSIS OF VARIANCE

SOURCE	SUM-OF-SQUARES	DF	MEAN-SQUARE	F-RATIO	P
REGRESSION	3907.349	2	1953.674	77.308	0.003
RESIDUAL	75.814	3	25.271		

MODEL CONTAINS NO CONSTANT.

DEP VAR: Swamp C174 N: 6 MULTIPLE R: 0.996 SQUARED MULTIPLE R: 0.992
ADJUSTED SQUARED MULTIPLE R: 0.990 STANDARD ERROR OF ESTIMATE: 5.684

VARIABLE	COEFFICIENT	STD ERROR	STD COEF	TOLERANCE	T	P(2 TAIL)
Leaf C133	0.241	0.040	0.372	0.542	6.066	0.004
Water B190	0.753	0.065	0.706	0.542	11.508	0.000

ANALYSIS OF VARIANCE

SOURCE	SUM-OF-SQUARES	DF	MEAN-SQUARE	F-RATIO	P
REGRESSION	15732.427	2	7866.214	243.517	0.000
RESIDUAL	129.210	4	32.302		

DEP VAR: Swamp C174 N: 6 MULTIPLE R: 0.956 SQUARED MULTIPLE R: 0.915
ADJUSTED SQUARED MULTIPLE R: 0.858 STANDARD ERROR OF ESTIMATE: 10.642

VARIABLE	COEFFICIENT	STD ERROR	STD COEF	TOLERANCE	T	P(2 TAIL)
CONSTANT	-3.315	17.755	0.000	.	-0.187	0.864
Concrete B162	0.177	0.141	0.239	0.783	1.255	0.298
Water B190	0.662	0.154	0.821	0.783	4.310	0.023

ANALYSIS OF VARIANCE

SOURCE	SUM-OF-SQUARES	DF	MEAN-SQUARE	F-RATIO	P
REGRESSION	3643.416	2	1821.708	16.086	0.025
RESIDUAL	339.747	3	113.249		

Regression and ANOVA Tables Used in the Mixture Analysis (continued)

MODEL CONTAINS NO CONSTANT.

DEP VAR: Swamp C174 N: 6 MULTIPLE R: 0.989 SQUARED MULTIPLE R: 0.978
ADJUSTED SQUARED MULTIPLE R: 0.973 STANDARD ERROR OF ESTIMATE: 9.269

VARIABLE	COEFFICIENT	STD ERROR	STD COEF	TOLERANCE	T	P(2 TAIL)
Concrete B162	0.152	0.045	0.414	0.357	3.367	0.028
Water B190	0.667	0.131	0.625	0.357	5.080	0.007

ANALYSIS OF VARIANCE

SOURCE	SUM-OF-SQUARES	DF	MEAN-SQUARE	F-RATIO	P
REGRESSION	15517.943	2	7758.971	90.301	0.000
RESIDUAL	343.694	4	85.924		

MODEL CONTAINS NO CONSTANT.

DEP VAR: Swamp C174 N: 6 MULTIPLE R: 0.999 SQUARED MULTIPLE R: 0.999
ADJUSTED SQUARED MULTIPLE R: 0.999 STANDARD ERROR OF ESTIMATE: 2.066

VARIABLE	COEFFICIENT	STD ERROR	STD COEF	TOLERANCE	T	P(2 TAIL)
Grass C125	0.220	0.013	0.395	0.525	17.457	0.000
Water B190	0.731	0.024	0.685	0.525	30.274	0.000

ANALYSIS OF VARIANCE

SOURCE	SUM-OF-SQUARES	DF	MEAN-SQUARE	F-RATIO	P
REGRESSION	15844.564	2	7922.282	1856.111	0.000
RESIDUAL	17.073	4	4.268		

MODEL CONTAINS NO CONSTANT.

DEP VAR: Swamp C174 N: 6 MULTIPLE R: 0.998 SQUARED MULTIPLE R: 0.997
ADJUSTED SQUARED MULTIPLE R: 0.996 STANDARD ERROR OF ESTIMATE: 3.508

VARIABLE	COEFFICIENT	STD ERROR	STD COEF	TOLERANCE	T	P(2 TAIL)
Grass C123	0.232	0.023	0.418	0.457	10.154	0.001
Water B190	0.693	0.044	0.649	0.457	15.755	0.000

ANALYSIS OF VARIANCE

SOURCE	SUM-OF-SQUARES	DF	MEAN-SQUARE	F-RATIO	P
REGRESSION	15812.424	2	7906.212	642.604	0.000
RESIDUAL	49.214	4	12.303		

Regression and ANOVA Tables Used in the Mixture Analysis (continued)

MODEL CONTAINS NO CONSTANT.

DEP VAR: Swamp C174   N: 6   MULTIPLE R: 0.999   SQUARED MULTIPLE R: 0.997
ADJUSTED SQUARED MULTIPLE R: 0.997   STANDARD ERROR OF ESTIMATE: 3.249

VARIABLE	COEFFICIENT	STD ERROR	STD COEF	TOLERANCE	T	P(2 TAIL)
Pine B140	0.402	0.037	0.492	0.332	10.993	0.000
Water B190	0.593	0.048	0.555	0.332	12.398	0.000

ANALYSIS OF VARIANCE

SOURCE	SUM-OF-SQUARES	DF	MEAN-SQUARE	F-RATIO	P
REGRESSION	15819.416	2	7909.708	749.357	0.000
RESIDUAL	42.221	4	10.555		

MODEL CONTAINS NO CONSTANT.

DEP VAR: Swamp C174   N: 6   MULTIPLE R: 0.996   SQUARED MULTIPLE R: 0.992
ADJUSTED SQUARED MULTIPLE R: 0.989   STANDARD ERROR OF ESTIMATE: 5.780

VARIABLE	COEFFICIENT	STD ERROR	STD COEF	TOLERANCE	T	P(2 TAIL)
Asphalt B160	0.572	0.084	0.740	0.176	6.770	0.002
Pine B140	0.224	0.089	0.275	0.176	2.516	0.066

ANALYSIS OF VARIANCE

SOURCE	SUM-OF-SQUARES	DF	MEAN-SQUARE	F-RATIO	P
REGRESSION	15728.014	2	7864.007	235.409	0.000
RESIDUAL	133.623	4	33.406		

Regression and ANOVA Tables Used in the Mixture Analysis (continued)

MODEL CONTAINS NO CONSTANT.

DEP VAR: Swamp C174 N: 6 MULTIPLE R: 0.989 SQUARED MULTIPLE R: 0.979
ADJUSTED SQUARED MULTIPLE R: 0.974 STANDARD ERROR OF ESTIMATE: 9.150

VARIABLE	COEFFICIENT	STD ERROR	STD COEF	TOLERANCE	T	P(2 TAIL)
Asphalt B160	0.697	0.203	0.901	0.076	3.426	0.027
Water B190	0.098	0.281	0.092	0.076	0.348	0.745

ANALYSIS OF VARIANCE

SOURCE	SUM-OF-SQUARES	DF	MEAN-SQUARE	F-RATIO	P
REGRESSION	15526.728	2	7763.364	92.722	0.000
RESIDUAL	334.909	4	83.727		

MODEL CONTAINS NO CONSTANT.

DEP VAR: Swamp C174 N: 6 MULTIPLE R: 0.996 SQUARED MULTIPLE R: 0.991
ADJUSTED SQUARED MULTIPLE R: 0.989 STANDARD ERROR OF ESTIMATE: 5.875

VARIABLE	COEFFICIENT	STD ERROR	STD COEF	TOLERANCE	T	P(2 TAIL)
Asphalt B160	0.652	0.059	0.843	0.379	11.122	0.000
Leaf C133	0.120	0.049	0.186	0.379	2.449	0.071

ANALYSIS OF VARIANCE

SOURCE	SUM-OF-SQUARES	DF	MEAN-SQUARE	F-RATIO	P
REGRESSION	15723.595	2	7861.798	227.809	0.000
RESIDUAL	138.042	4	34.510		

MODEL CONTAINS NO CONSTANT.

DEP VAR: Swamp C174 N: 6 MULTIPLE R: 0.993 SQUARED MULTIPLE R: 0.987
ADJUSTED SQUARED MULTIPLE R: 0.983 STANDARD ERROR OF ESTIMATE: 7.269

VARIABLE	COEFFICIENT	STD ERROR	STD COEF	TOLERANCE	T	P(2 TAIL)
Grass C125	0.090	0.057	0.162	0.320	1.591	0.187
Asphalt B160	0.661	0.079	0.855	0.320	8.388	0.001

ANALYSIS OF VARIANCE

SOURCE	SUM-OF-SQUARES	DF	MEAN-SQUARE	F-RATIO	P
REGRESSION	15650.270	2	7825.135	148.086	0.000
RESIDUAL	211.367	4	52.842		

Regression and ANOVA Tables Used in the Mixture Analysis (continued)

MODEL CONTAINS NO CONSTANT.

DEP VAR: Swamp C176   N: 6   MULTIPLE R: 0.999   SQUARED MULTIPLE R: 0.999
ADJUSTED SQUARED MULTIPLE R: 0.999   STANDARD ERROR OF ESTIMATE: 1.974

VARIABLE	COEFFICIENT	STD ERROR	STD COEF	TOLERANCE	T	P(2 TAIL)
Leaf C133	0.134	0.014	0.223	0.542	9.693	0.001
Water B190	0.827	0.023	0.835	0.542	36.357	0.000

ANALYSIS OF VARIANCE

SOURCE	SUM-OF-SQUARES	DF	MEAN-SQUARE	F-RATIO	P
REGRESSION	13622.411	2	6811.205	1747.836	0.000
RESIDUAL	15.588	4	3.897		

MODEL CONTAINS NO CONSTANT.

DEP VAR: Swamp C176   N: 6   MULTIPLE R: 0.995   SQUARED MULTIPLE R: 0.990
ADJUSTED SQUARED MULTIPLE R: 0.987   STANDARD ERROR OF ESTIMATE: 5.846

VARIABLE	COEFFICIENT	STD ERROR	STD COEF	TOLERANCE	T	P(2 TAIL)
Concrete B162	0.076	0.029	0.224	0.357	2.677	0.055
Water B190	0.798	0.083	0.806	0.357	9.629	0.001

ANALYSIS OF VARIANCE

SOURCE	SUM-OF-SQUARES	DF	MEAN-SQUARE	F-RATIO	P
REGRESSION	13501.276	2	6750.638	197.498	0.000
RESIDUAL	136.723	4	34.181		

MODEL CONTAINS NO CONSTANT.

DEP VAR: Swamp C176   N: 6   MULTIPLE R: 1.000   SQUARED MULTIPLE R: 1.000
ADJUSTED SQUARED MULTIPLE R: 1.000   STANDARD ERROR OF ESTIMATE: 0.959

VARIABLE	COEFFICIENT	STD ERROR	STD COEF	TOLERANCE	T	P(2 TAIL)
Grass C125	0.119	0.006	0.230	0.525	20.282	0.000
Water B190	0.819	0.011	0.828	0.525	73.060	0.000

ANALYSIS OF VARIANCE

SOURCE	SUM-OF-SQUARES	DF	MEAN-SQUARE	F-RATIO	P
REGRESSION	13634.322	2	681[illegible].161	[illegible]417.773	0.000
RESIDUAL	3.676	4	0.9[illegible]9		

Regression and ANOVA Tables Used in the Mixture Analysis (continued)

MODEL CONTAINS NO CONSTANT.

DEP VAR: Swamp C176 N: 6 MULTIPLE R: 0.985 SQUARED MULTIPLE R: 0.970
ADJUSTED SQUARED MULTIPLE R: 0.962 STANDARD ERROR OF ESTIMATE: 10.143

VARIABLE	COEFFICIENT	STD ERROR	STD COEF	TOLERANCE	T	P(2 TAIL)
Asphalt B160	0.727	0.110	1.015	0.320	6.612	0.003
Grass C125	-0.019	0.079	-0.036	0.320	-0.238	0.824

ANALYSIS OF VARIANCE

SOURCE	SUM-OF-SQUARES	DF	MEAN-SQUARE	F-RATIO	P
REGRESSION	13226.475	2	6613.237	64.281	0.001
RESIDUAL	411.524	4	102.881		

MODEL CONTAINS NO CONSTANT.

DEP VAR: Swamp C176 N: 6 MULTIPLE R: 1.000 SQUARED MULTIPLE R: 1.000
ADJUSTED SQUARED MULTIPLE R: 1.000 STANDARD ERROR OF ESTIMATE: 0.978

VARIABLE	COEFFICIENT	STD ERROR	STD COEF	TOLERANCE	T	P(2 TAIL)
Pine B140	0.219	0.011	0.289	0.332	19.871	0.000
Water B190	0.742	0.014	0.750	0.332	51.555	0.000

ANALYSIS OF VARIANCE

SOURCE	SUM-OF-SQUARES	DF	MEAN-SQUARE	F-RATIO	P
REGRESSION	13634.170	2	6817.085	7123.011	0.000
RESIDUAL	3.828	4	0.957		

DEP VAR: Swamp C176 N: 6 MULTIPLE R: 0.979 SQUARED MULTIPLE R: 0.958
ADJUSTED SQUARED MULTIPLE R: 0.931 STANDARD ERROR OF ESTIMATE: 7.867

VARIABLE	COEFFICIENT	STD ERROR	STD COEF	TOLERANCE	T	P(2 TAIL)
CONSTANT	-14.536	7.601	0.000	.	-1.912	0.152
Pine B140	0.145	0.134	0.155	0.672	1.082	0.358
Asphalt B160	0.774	0.126	0.882	0.672	6.142	0.009

ANALYSIS OF VARIANCE

SOURCE	SUM-OF-SQUARES	DF	MEAN-SQUARE	F-RATIO	P
REGRESSION	4285.012	2	2142.506	34.616	0.008
RESIDUAL	185.678	3	61.893		

Regression and ANOVA Tables Used in the Mixture Analysis (continued)

DEP VAR: <u>Swamp C176</u> N: 6 MULTIPLE R: 0.971 SQUARED MULTIPLE R: 0.942
ADJUSTED SQUARED MULTIPLE R: 0.928 STANDARD ERROR OF ESTIMATE: 8.034

VARIABLE	COEFFICIENT	STD ERROR	STD COEF	TOLERANCE	T	P(2 TAIL)
CONSTANT	-11.018	7.016	0.000	.	-1.570	0.191
Asphalt B160	0.852	0.106	0.971	1.000	8.079	0.001

ANALYSIS OF VARIANCE

SOURCE	SUM-OF-SQUARES	DF	MEAN-SQUARE	F-RATIO	P
REGRESSION	4212.539	1	4212.539	<u>65.272</u>	0.001
RESIDUAL	258.151	4	64.538		

MODEL CONTAINS NO CONSTANT.

DEP VAR: <u>Swamp C176</u> N: 6 MULTIPLE R: 0.985 SQUARED MULTIPLE R: 0.970
ADJUSTED SQUARED MULTIPLE R: 0.962 STANDARD ERROR OF ESTIMATE: 10.149

VARIABLE	COEFFICIENT	STD ERROR	STD COEF	TOLERANCE	T	P(2 TAIL)
Pine B140	0.035	0.157	0.047	0.176	0.226	0.832
Asphalt B160	0.675	0.148	0.942	0.176	4.553	0.010

ANALYSIS OF VARIANCE

SOURCE	SUM-OF-SQUARES	DF	MEAN-SQUARE	F-RATIO	P
REGRESSION	13225.952	2	6612.976	<u>64.196</u>	0.001
RESIDUAL	412.047	4	103.012		

Regression and ANOVA Tables Used in the Mixture Analysis (continued)

MODEL CONTAINS NO CONSTANT.

DEP VAR: Swamp C175 N: 6 MULTIPLE R: 0.990 SQUARED MULTIPLE R: 0.980
ADJUSTED SQUARED MULTIPLE R: 0.975 STANDARD ERROR OF ESTIMATE: 9.050

VARIABLE	COEFFICIENT	STD ERROR	STD COEF	TOLERANCE	T	P(2 TAIL)
Leaf C133	0.280	0.063	0.423	0.542	4.428	0.011
Water B190	0.713	0.104	0.654	0.542	6.844	0.002

ANALYSIS OF VARIANCE

SOURCE	SUM-OF-SQUARES	DF	MEAN-SQUARE	F-RATIO	P
REGRESSION	16256.170	2	8128.085	99.243	0.000
RESIDUAL	327.604	4	81.901		

MODEL CONTAINS NO CONSTANT.

DEP VAR: Swamp C175 N: 6 MULTIPLE R: 0.997 SQUARED MULTIPLE R: 0.995
ADJUSTED SQUARED MULTIPLE R: 0.994 STANDARD ERROR OF ESTIMATE: 4.625

VARIABLE	COEFFICIENT	STD ERROR	STD COEF	TOLERANCE	T	P(2 TAIL)
Grass C125	0.263	0.028	0.461	0.525	9.295	0.001
Water B190	0.679	0.054	0.622	0.525	12.561	0.000

ANALYSIS OF VARIANCE

SOURCE	SUM-OF-SQUARES	DF	MEAN-SQUARE	F-RATIO	P
REGRESSION	16498.212	2	8249.106	385.647	0.000
RESIDUAL	85.561	4	21.390		

DEP VAR: Swamp C175 N: 6 MULTIPLE R: 0.999 SQUARED MULTIPLE R: 0.998
ADJUSTED SQUARED MULTIPLE R: 0.996 STANDARD ERROR OF ESTIMATE: 1.642

VARIABLE	COEFFICIENT	STD ERROR	STD COEF	TOLERANCE	T	P(2 TAIL)
CONSTANT	-5.700	1.616	0.000	.	-3.527	0.039
Leaf C133	0.175	0.015	0.319	0.956	11.354	0.001
Asphalt B160	0.693	0.022	0.882	0.956	31.400	0.000

ANALYSIS OF VARIANCE

SOURCE	SUM-OF-SQUARES	DF	MEAN-SQUARE	F-RATIO	P
REGRESSION	3566.358	2	1783.179	661.567	0.000
RESIDUAL	8.086	3	2.695		

Regression and ANOVA Tables Used in the Mixture Analysis (continued)

MODEL CONTAINS NO CONSTANT.

DEP VAR: Swamp C175 N: 6 MULTIPLE R: 0.999 SQUARED MULTIPLE R: 0.997
ADJUSTED SQUARED MULTIPLE R: 0.997 STANDARD ERROR OF ESTIMATE: 3.226

VARIABLE	COEFFICIENT	STD ERROR	STD COEF	TOLERANCE	T	P(2 TAIL)
Leaf C133	0.150	0.027	0.227	0.379	5.580	0.005
Asphalt B160	0.640	0.032	0.810	0.379	19.903	0.000

ANALYSIS OF VARIANCE

SOURCE	SUM-OF-SQUARES	DF	MEAN-SQUARE	F-RATIO	P
REGRESSION	16542.149	2	8271.074	794.824	0.000
RESIDUAL	41.625	4	10.406		

MODEL CONTAINS NO CONSTANT.

DEP VAR: Swamp C175 N: 6 MULTIPLE R: 0.998 SQUARED MULTIPLE R: 0.995
ADJUSTED SQUARED MULTIPLE R: 0.994 STANDARD ERROR OF ESTIMATE: 4.467

VARIABLE	COEFFICIENT	STD ERROR	STD COEF	TOLERANCE	T	P(2 TAIL)
Grass C125	0.132	0.035	0.232	0.320	3.785	0.019
Asphalt B160	0.631	0.048	0.798	0.320	13.017	0.000

ANALYSIS OF VARIANCE

SOURCE	SUM-OF-SQUARES	DF	MEAN-SQUARE	F-RATIO	P
REGRESSION	16503.964	2	8251.982	413.586	0.000
RESIDUAL	79.809	4	19.952		

MODEL CONTAINS NO CONSTANT.

DEP VAR: Swamp C175 N: 6 MULTIPLE R: 0.990 SQUARED MULTIPLE R: 0.979
ADJUSTED SQUARED MULTIPLE R: 0.974 STANDARD ERROR OF ESTIMATE: 9.239

VARIABLE	COEFFICIENT	STD ERROR	STD COEF	TOLERANCE	T	P(2 TAIL)
Water B190	-0.151	0.283	-0.138	0.076	-0.531	0.623
Asphalt B160	0.887	0.205	1.122	0.076	4.318	0.012

ANALYSIS OF VARIANCE

SOURCE	SUM-OF-SQUARES	DF	MEAN-SQUARE	F-RATIO	P
REGRESSION	16242.305	2	8121.153	95.132	0.000
RESIDUAL	341.468	4	85.367		

APPENDIX E: Linear Model Results for Three Endmembers

Regression Results for Three-Endmember Mixture Analysis

DEP VAR: [illegible] C174 N: 6 MULTIPLE R: 0.999 SQUARED MULTIPLE R: 0.999
ADJUSTED SQUARED MULTIPLE R: 0.998 STANDARD ERROR OF ESTIMATE: 2.459

VARIABLE	COEFFICIENT	STD ERROR	STD COEF	TOLERANCE	T	P(2 TAIL)
Leaf C133	0.172	0.023	0.266	0.290	7.338	0.005
Concrete B162	0.070	0.016	0.191	0.191	4.286	0.023
Water B190	0.666	0.035	0.624	0.357	19.113	0.000

ANALYSIS OF VARIANCE

SOURCE	SUM-OF-SQUARES	DF	MEAN-SQUARE	F-RATIO	P
REGRESSION	15843.500	3	5281.167	873.522	0.000
RESIDUAL	18.137	3	6.046		

RESIDUALS

	B1	B2	B3	B4	B5	B7
Leaf, Water	-1.05	-1.28	2.68	-4.77	8.64	4.68
Concrete,Water	2.01	-4.28	-8.29	14.71	0.63	-5.97
Concrete,Water, Leaf	1.45	-2.30	-1.93	-1.11	2.30	-0.71

DURBIN-WATSON D STATISTIC 1.961
FIRST ORDER AUTOCORRELATION -0.052

EIGENVALUES OF UNIT SCALED X'X

1	2	3
2.550	0.325	0.125

CONDITION INDICES

1	2	3
1.000	2.799	4.518

VARIANCE PROPORTIONS

	1	2	3
C133	0.037	0.375	0.588
B162	0.027	0.005	0.968
B190	0.044	0.628	0.328

CORRELATION MATRIX OF REGRESSION COEFFICIENTS

	C133	B162	B190
C133	1.000		
B162	-0.682	1.000	
B190	-0.005	-0.583	1.000

Regression Results for Three-Endmember Mixture Analysis (continued)

DEP VAR: Swamp C174 N: 6 MULTIPLE R: 1.000 SQUARED MULTIPLE R: 1.000
ADJUSTED SQUARED MULTIPLE R: 0.999 STANDARD ERROR OF ESTIMATE: 1.361

VARIABLE	COEFFICIENT	STD ERROR	STD COEF	TOLERANCE	T	P(2 TAIL)
Grass C125	0.192	0.014	0.344	0.[illegible]30	13.512	0.001
Concrete B162	0.028	0.011	0.077	0.123	2.494	0.088
Water B190	0.703	0.019	0.659	0.351	36.125	0.000

ANALYSIS OF VARIANCE

SOURCE	SUM-OF-SQUARES	DF	MEAN-SQUARE	F-RATIO	P
REGRESSION	15856.081	3	5285.360	2853.978	0.000
RESIDUAL	5.556	3	1.852		

RESIDUALS

	B1	B2	B3	B4	B5	B7
Concrete,Water, Grass	0.77	-2.00	-0.08	-0.49	0.80	-0.25

DURBIN-WATSON D STATISTIC 2.575
FIRST ORDER AUTOCORRELATION -0.346

EIGENVALUES OF UNIT SCALED X'X

1	2	3
2.599	0.324	0.077

CONDITION INDICES

1	2	3
1.000	2.831	5.828

VARIANCE PROPORTIONS

	1	2	3
C125	0.023	0.184	0.793
B162	0.017	0.012	0.971
B190	0.041	0.690	0.269

CORRELATION MATRIX OF REGRESSION COEFFICIENTS

	C125	B162	B190
C125	1.000		
B162	-0.810	1.000	
B190	0.137	-0.576	1.000

LIST OF ACRONYMS

AFB	Air Force Base
AVIRIS	Airborne Visible/ Infrared Imaging Spectrometer
DMA	Defense Mapping Agency
DOC	Degree of Compliance
GSD	Ground Sampling Distance
GT	Ground Truth
IFOV	Instantaneous Field of View
ISODATA	Iterative Self-Organizing Data Analysis Techniques A
JPL	Jet Propulsion Laboratory
LAS	Land Analysis System
MVN	Multivariate Normal
NHAP	National High Altitude Photography
RGB	Red, Green, Blue
RW	Runway
SPL	TEC's Space Programs Laboratory
SRTF/MBIPS	Space Research Test Facility, Multiband Image Processing System
TEC	U.S. Army Topographic Engineer Center
TM	Landsat Thematic Mapper
TTADB	DMA's Tactical Terrain Analysis Data Base

END

FILMED

DATE:

1-94

DTIC